Best unbiased estimators and sufficiency

Sahir Rai Bhatnagar
Department of Epidemiology, Biostatistics, and Occupational Health
Department of Diagnostic Radiology
McGill University

sahirbhatnagar.com

November 15, 2021

Recap
- Point Estimation
- Maximum Likelihood Estimator
- MLE example

Methods of Evaluating Estimators
- Uniformly Minimum Variance Unbiased Estimator
- Challenges in finding the UMVUE
- Cramér–Rao Lower Bound (CRLB)
- Fisher Information
- Correlation review
- CRLB proof
- Attainment of CRLB

Rao-Blackwell

Recap
Point Estimation
Maximum Likelihood Estimator
MLE example

Methods of Evaluating Estimators
Uniformly Minimum Variance Unbiased Estimator
Challenges in finding the UMVUE
Cramér–Rao Lower Bound (CRLB)
Fisher Information
Correlation review
CRLB proof
Attainment of CRLB

Rao-Blackwell

Point Estimation

- When sampling is from a population described by a pdf or a pmf $f(x|\theta)$, knowledge of θ yields knowledge of the entire population.

Point Estimation

- When sampling is from a population described by a pdf or a pmf $f(x|\theta)$, knowledge of θ yields knowledge of the entire population.
- Hence it is natural to seek a method of finding a good estimator of the point θ, i.e., a good *point estimator*

Point Estimation

- When sampling is from a population described by a pdf or a pmf $f(x|\theta)$, knowledge of θ yields knowledge of the entire population.
- Hence it is natural to seek a method of finding a good estimator of the point θ, i.e., a good *point estimator*

Definition 1.1 (Point estimator).

A *point estimator* is any function $W(X_1, \ldots, X_n)$ of a sample; that is, any statistic is a point estimator.

Point Estimation

- When sampling is from a population described by a pdf or a pmf $f(x|\theta)$, knowledge of θ yields knowledge of the entire population.
- Hence it is natural to seek a method of finding a good estimator of the point θ, i.e., a good *point estimator*

Definition 1.1 (Point estimator).

A *point estimator* is any function $W(X_1, \ldots, X_n)$ of a sample; that is, any statistic is a point estimator.

- An *estimator* is a function of the sample, while an *estimate* is the realized value of an estimator (a number) that is obtained when a sample is actually taken

Point Estimation

- When sampling is from a population described by a pdf or a pmf $f(x|\theta)$, knowledge of θ yields knowledge of the entire population.
- Hence it is natural to seek a method of finding a good estimator of the point θ, i.e., a good *point estimator*

Definition 1.1 (Point estimator).

A *point estimator* is any function $W(X_1, \ldots, X_n)$ of a sample; that is, any statistic is a point estimator.

- An *estimator* is a function of the sample, while an *estimate* is the realized value of an estimator (a number) that is obtained when a sample is actually taken
- Notationally, when a sample is taken, an estimator is a function of the random variables $X_1, \ldots, X_n$, while an estimate is a function of the realized values $x_1, \ldots, x_n$.

Methods of Finding Estimators

- In some cases, there will be an obvious or natural candidate for a point estimator of a particular parameter, e.g., the sample mean ($\bar{X}$) as a point estimator of the population mean (μ)

Methods of Finding Estimators

- In some cases, there will be an obvious or natural candidate for a point estimator of a particular parameter, e.g., the sample mean ($\bar{X}$) as a point estimator of the population mean (μ)
- However, when we leave a simple case like this, intuition may not only desert us, it may also lead us astray. Therefore, it is useful to have some techniques that will at least give us some reasonable candidates for consideration.

Methods of Finding Estimators

- In some cases, there will be an obvious or natural candidate for a point estimator of a particular parameter, e.g., the sample mean ($\bar{X}$) as a point estimator of the population mean (μ)
- However, when we leave a simple case like this, intuition may not only desert us, it may also lead us astray. Therefore, it is useful to have some techniques that will at least give us some reasonable candidates for consideration.
- Maximum Likelihood Estimators is, by far, the most popular technique for deriving estimators

Maximum Likelihood Estimator

- Let $X_1, \ldots, X_n$ be an iid sample from a population with pdf or pmf $f(x|\Theta)$ where $\Theta \equiv (\theta_1, \ldots, \theta_k)$ have unknown values and $\mathbf{x} = x_1, \ldots, x_n$ are the observed sample values. The likelihood function is defined by

$$L(\Theta|\mathbf{x}) = L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k) \tag{1}$$

Definition 1.2 (Maximum Likelihood Estimator).

For each sample point $\mathbf{x}$, let $\widehat{\Theta}$ be a parameter value at which $L(\Theta|\mathbf{x})$ attains its maximum as a function of Θ, with $\mathbf{x}$ held fixed. A *maximum likelihood estimator (MLE)* of the parameter Θ based on a sample $\mathbf{X}$ is $\widehat{\Theta}(\mathbf{X})$.

- $\widehat{\Theta}(\mathbf{x})$ is called the maximum likelihood estimate of Θ based on data $\mathbf{x}$
- $\widehat{\Theta}(\mathbf{X})$ is the maximum likelihood estimator (MLE) of Θ

Example 1 (Poisson MLE).

Let $X_1, \ldots, X_n$ be iid Poisson(λ). Then the likelihood function is

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \tag{2}$$

The log-likehood is given by

$$\ell(\lambda|\mathbf{x}) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log x_i! \tag{3}$$

Example 1 (Poisson MLE).

Let $X_1, \ldots, X_n$ be iid Poisson(λ). Then the likelihood function is

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \tag{2}$$

The log-likehood is given by

$$\ell(\lambda|\mathbf{x}) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log x_i! \tag{3}$$

Taking the derivative of (3) with respect to λ we get:

$$\frac{\partial \ell(\lambda|\mathbf{x})}{\partial \lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \tag{4}$$

Example 1 (Poisson MLE).

Let $X_1, \ldots, X_n$ be iid Poisson(λ). Then the likelihood function is

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \tag{2}$$

The log-likehood is given by

$$\ell(\lambda|\mathbf{x}) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log x_i! \tag{3}$$

Taking the derivative of (3) with respect to λ we get:

$$\frac{\partial \ell(\lambda|\mathbf{x})}{\partial \lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \tag{4}$$

Setting (4) equal to zero for the first order condition, and solving for λ, yields $\hat{\lambda} = \bar{x}$.

Example 1 (Poisson MLE).

Let $X_1, \ldots, X_n$ be iid Poisson(λ). Then the likelihood function is

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \tag{2}$$

The log-likehood is given by

$$\ell(\lambda|\mathbf{x}) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log x_i! \tag{3}$$

Taking the derivative of (3) with respect to λ we get:

$$\frac{\partial \ell(\lambda|\mathbf{x})}{\partial \lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \tag{4}$$

Setting (4) equal to zero for the first order condition, and solving for λ, yields $\hat{\lambda} = \bar{x}$. In order to verify that this is the MLE for λ we take the second derivative of (3) with respect to λ:

$$\frac{\partial^2 \ell(\lambda|\mathbf{x})}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} \tag{5}$$

Since (5) is negative and the log-likelihood is concave, $\hat{\lambda} = \bar{x}$ solves for the global maximum.

Poisson Unbiased Estimation

Theorem 1.3 (Relationships between a statistic and population parameter).

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$.

$$\begin{aligned} &a.\ \mathrm{E}\bar{X} = \mu \\ &b.\ \mathrm{Var}\,\bar{X} = \frac{\sigma^2}{n} \\ &c.\ \mathrm{E}S^2 = \sigma^2 \end{aligned}$$

where $\bar{X}$ and S^2 are the sample mean and sample variance, respectively.

Poisson Unbiased Estimation

Theorem 1.3 (Relationships between a statistic and population parameter).

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$.

$$\begin{aligned} &a.\ \mathrm{E}\bar{X} = \mu \\ &b.\ \operatorname{Var}\bar{X} = \frac{\sigma^2}{n} \\ &c.\ \mathrm{E}S^2 = \sigma^2 \end{aligned}$$

where $\bar{X}$ and S^2 are the sample mean and sample variance, respectively.

Applying Theorem 1.3 to $X_1, \ldots, X_n$ iid Poisson(λ) we have

$$\mathrm{E}_\lambda \bar{X} = \lambda, \qquad \text{for all } \lambda,$$
$$\mathrm{E}_\lambda S^2 = \lambda, \qquad \text{for all } \lambda,$$

so both $\bar{X}$ and S^2 are unbiased estimators of λ.

Poisson MLE using optim

We can use the `stats::optim` function in `R` to find the MLE, provided we have a likelihood function. The `optim` can maximize (or minimize) an objective function using many different algorithms. This is referred to as **solving the objective function numerically**.

```r
# define the objective function
ll.poisson <- function(lambda, x) {
  sum(x) * log(lambda) - length(x) * lambda
}

# generate some data
data <- rpois(1e6, 5)

# by default optim finds the min, but the
# negative min is the max therefore we need
# to use list(fnscale = -1)
opt <- optim(par = 2,
             fn = ll.poisson,
             method = "BFGS",
             control = list(fnscale = -1),
             x = data)

# compare numeric vs. analytical solutions
c(numerical = opt$par,
  xbar = mean(data),
  samplevar = var(data))
```

```
## numerical      xbar samplevar
##  5.000532  5.000535  5.015680
```

```r
curve(ll.poisson(x, data), 0,10, xlab = "lambda",lwd = 4)
abline(h = opt$value, v = opt$par, lty = 2, lwd = 4, col = "red")
```

Recap
Point Estimation
Maximum Likelihood Estimator
MLE example

Methods of Evaluating Estimators
Uniformly Minimum Variance Unbiased Estimator
Challenges in finding the UMVUE
Cramér–Rao Lower Bound (CRLB)
Fisher Information
Correlation review
CRLB proof
Attainment of CRLB

Rao-Blackwell

Methods of Evaluating Estimators

- As we saw in the Poisson example, $\bar{X}$ and S^2 are both unbiased estimators of λ. *How do we choose between these estimators?*

Methods of Evaluating Estimators

- As we saw in the Poisson example, $\bar{X}$ and S^2 are both unbiased estimators of λ. *How do we choose between these estimators?*
- The mean squared error (MSE) of an estimator W of a parameter θ is the function of θ defined by

$$\mathrm{E}_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + (Bias_\theta W)^2$$

Methods of Evaluating Estimators

- As we saw in the Poisson example, $\bar{X}$ and S^2 are both unbiased estimators of λ. *How do we choose between these estimators?*
- The mean squared error (MSE) of an estimator W of a parameter θ is the function of θ defined by

$$\mathrm{E}_\theta (W - \theta)^2 = \mathrm{Var}_\theta W + (Bias_\theta W)^2$$

- If W_1 and W_2 are both unbiased estimators of a parameter θ, that is, $\mathrm{E}_\theta W_1 = \mathrm{E}_\theta W_2 = \theta$, then their MSE is equal to their variances $\rightarrow$ we should choose the estimator with the smaller variance

Methods of Evaluating Estimators

- As we saw in the Poisson example, $\bar{X}$ and S^2 are both unbiased estimators of λ. *How do we choose between these estimators?*
- The mean squared error (MSE) of an estimator W of a parameter θ is the function of θ defined by

$$\mathrm{E}_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + (Bias_\theta W)^2$$

- If W_1 and W_2 are both unbiased estimators of a parameter θ, that is, $\mathrm{E}_\theta W_1 = \mathrm{E}_\theta W_2 = \theta$, then their MSE is equal to their variances $\rightarrow$ we should choose the estimator with the smaller variance
- If we can find an unbiased estimator with uniformly smallest variance – a best unbiased estimator – then our task is done.

Uniformly Minimum Variance Unbiased Estimator

- The goal of this section is to investigate a method for finding a "best" unbiased estimator defined in the following way:

Definition 1.4 (UMVUE).

$W^*(\mathbf{X})$ is the best unbiased estimator, or uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ if

1. $\mathrm{E}_\theta\left[W^*(\mathbf{X}) \mid \theta\right] = \tau(\theta)$ for all θ (unbiased)
2. $\mathrm{Var}\left[W^*(\mathbf{X}) \mid \theta\right] \leq \mathrm{Var}[W(\mathbf{X}) \mid \theta]$ for all θ, where W is any other unbiased estimator of $\tau(\theta)$ (minimum variance).

Uniformly Minimum Variance Unbiased Estimator

- The goal of this section is to investigate a method for finding a "best" unbiased estimator defined in the following way:

Definition 1.4 (UMVUE).

$W^*(\mathbf{X})$ is the best unbiased estimator, or uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ if

1. $\mathrm{E}_\theta\left[W^*(\mathbf{X}) \mid \theta\right] = \tau(\theta)$ for all θ (unbiased)
2. $\mathrm{Var}\left[W^*(\mathbf{X}) \mid \theta\right] \leq \mathrm{Var}[W(\mathbf{X}) \mid \theta]$ for all θ, where W is any other unbiased estimator of $\tau(\theta)$ (minimum variance).

- Finding a best unbiased estimator (if one exists) is not an easy task as we'll see in the next example

Poisson Unbiased Estimation Revisited

- Recall that by applying Theorem 1.3 to $X_1, \ldots, X_n$ iid Poisson(λ)

$$\begin{aligned} \mathrm{E}_\lambda \bar{X} &= \lambda, \qquad \text{for all } \lambda, \\ \mathrm{E}_\lambda S^2 &= \lambda, \qquad \text{for all } \lambda, \end{aligned}$$

  so both $\bar{X}$ and S^2 are unbiased estimators of λ.

Poisson Unbiased Estimation Revisited

- Recall that by applying Theorem 1.3 to $X_1, \ldots, X_n$ iid Poisson(λ)

$$\begin{aligned} \mathrm{E}_\lambda \bar{X} &= \lambda, \qquad \text{for all } \lambda, \\ \mathrm{E}_\lambda S^2 &= \lambda, \qquad \text{for all } \lambda, \end{aligned}$$

  so both $\bar{X}$ and S^2 are unbiased estimators of λ.
- Again from Theorem 1.3, we have $\mathrm{Var}_\lambda \bar{X} = \lambda/n$

Poisson Unbiased Estimation Revisited

- Recall that by applying Theorem 1.3 to $X_1, \ldots, X_n$ iid Poisson(λ)

$$\begin{aligned} \mathrm{E}_\lambda \bar{X} &= \lambda, \qquad && \text{for all } \lambda, \\ \mathrm{E}_\lambda S^2 &= \lambda, \qquad && \text{for all } \lambda, \end{aligned}$$

  so both $\bar{X}$ and S^2 are unbiased estimators of λ.
- Again from Theorem 1.3, we have $\mathrm{Var}_\lambda \bar{X} = \lambda / n$
- $\mathrm{Var}_\lambda \left[S^2\right] = \frac{1}{n}\mu_4 + \frac{\mu_2^2(n-3)}{n(n-1)}$ where μ_j is the j th centered moment $\rightarrow$ lengthy calculation

Poisson Unbiased Estimation Revisited

- Even if we can establish that $\bar{X}$ is better than S^2, consider the class of estimators
$$W_a(\bar{X}, S^2) = a\bar{X} + (1-a)S^2.$$
For every constant a, $\mathrm{E}_\lambda W_a(\bar{X}, S^2) = \lambda$, so we now have infinitely many unbiased estimators of λ.

Poisson Unbiased Estimation Revisited

- Even if we can establish that $\bar{X}$ is better than S^2, consider the class of estimators
$$W_a(\bar{X}, S^2) = a\bar{X} + (1-a)S^2.$$
For every constant a, $\mathrm{E}_\lambda W_a(\bar{X}, S^2) = \lambda$, so we now have infinitely many unbiased estimators of λ.

- Even if $\bar{X}$ is better than S^2, is it better than every $W_a(\bar{X}, S^2)$?

Poisson Unbiased Estimation Revisited

- Even if we can establish that $\bar{X}$ is better than S^2, consider the class of estimators
$$W_a(\bar{X}, S^2) = a\bar{X} + (1-a)S^2.$$
For every constant a, $\mathrm{E}_\lambda W_a(\bar{X}, S^2) = \lambda$, so we now have infinitely many unbiased estimators of λ.

- Even if $\bar{X}$ is better than S^2, is it better than every $W_a(\bar{X}, S^2)$?
- Furthermore, how can we be sure that there are not other, better, unbiased estimators lurking about?

Poisson Unbiased Estimation Revisited

- Even if we can establish that $\bar{X}$ is better than S^2, consider the class of estimators
$$W_a(\bar{X}, S^2) = a\bar{X} + (1-a)S^2.$$
For every constant a, $\mathrm{E}_\lambda W_a(\bar{X}, S^2) = \lambda$, so we now have infinitely many unbiased estimators of λ.

- Even if $\bar{X}$ is better than S^2, is it better than every $W_a(\bar{X}, S^2)$?
- Furthermore, how can we be sure that there are not other, better, unbiased estimators lurking about?
- This example shows some of the problems that might be encountered in trying to find a best unbiased estimator, and perhaps that a more comprehensive approach is desirable.

How to find the Best Unbiased Estimator?

- Find the lower bound of variances of any unbiased estimator of $\tau(\theta)$, say $B(\theta)$.
- If W^* is an unbiased estimator of $\tau(\theta)$ and satisfies $\mathrm{Var}\left[W^*(\mathbf{X}) \mid \theta\right] = B(\theta)$, then W^* is the best unbiased estimator.
- This is the appraoch taken with the use of the Cramér–Rao Lower Bound. The names Cramér and Rao are often interchanged depending on the textbook and professor's training.

Cramér–Rao Inequality

Theorem 1.5 (Cramér–Rao Lower Bound (CRLB)).

Let $X_1, \cdots, X_n$ *be iid with common pdf* $f(\mathbf{x} \mid \theta)$. *Let* $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ *be a statistic with mean* $\mathrm{E}_\theta W(\mathbf{X}) = k(\theta)$ *satisfying*

$$\frac{d}{d\theta}\mathrm{E}_\theta W(\mathbf{X}) = \frac{d}{d\theta}\int_{x \in \mathcal{X}} W(\mathbf{x}) f(\mathbf{x} \mid \theta)\, d\mathbf{x} = \int_{x \in \mathcal{X}} \frac{\partial}{\partial\theta} W(\mathbf{x}) f(\mathbf{x} \mid \theta)\, d\mathbf{x}$$

and

$$\mathrm{Var}_\theta\, W(\mathbf{X}) < \infty$$

Cramér–Rao Inequality

Theorem 1.5 (Cramér–Rao Lower Bound (CRLB)).

Let $X_1, \cdots, X_n$ *be iid with common pdf* $f(\mathbf{x} \mid \theta)$. *Let* $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ *be a statistic with mean* $\mathrm{E}_\theta W(\mathbf{X}) = k(\theta)$ *satisfying*

$$\frac{d}{d\theta}\mathrm{E}_\theta W(\mathbf{X}) = \frac{d}{d\theta}\int_{x\in\mathcal{X}} W(\mathbf{x})f(\mathbf{x} \mid \theta)\,d\mathbf{x} = \int_{x\in\mathcal{X}} \frac{\partial}{\partial\theta} W(\mathbf{x})f(\mathbf{x} \mid \theta)\,d\mathbf{x}$$

and

$$\mathrm{Var}_\theta\ W(\mathbf{X}) < \infty$$

Then, a lower bound of $\mathrm{Var}_\theta\ W(\mathbf{X})$ *is*

$$\mathrm{Var}_\theta\ W(\mathbf{X}) \geq \frac{\left[k'(\theta)\right]^2}{nI(\theta)}$$

where $I(\theta)$ *is the* ***Fisher information***

Fisher Information

- Before we prove the CRLB, we first recall 1) the Fisher information and 2) the formula for the covariance of two random variables.

Fisher Information

- Before we prove the CRLB, we first recall 1) the Fisher information and 2) the formula for the covariance of two random variables.

Theorem 1.6 (Fisher information).

If X is a random variable with pdf $f(x \mid \theta)$ which satisfies certain regularity assumptions then

$$\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right] = 0$$

and

$$\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right] = -\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right]$$

The quantity $I(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right]$ is called the information number or Fisher information.

Correlation review

- $\mathrm{E}[X] = \mu_X, \mathrm{E}[Y] = \mu_Y, \mathrm{Var}[X] = \sigma_X^2, \mathrm{Var}[Y] = \sigma_Y^2$
- Assume $0 < \sigma_X^2 < \infty$ and $0 < \sigma_Y^2 < \infty$

Definition 1.7 (Correlation coefficient).

The correlation of X and Y is the number defined by

$$\rho_{XY} = \frac{\mathrm{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

The value ρ_{XY} is also called the correlation coefficient.

Correlation review

- $\mathrm{E}[X] = \mu_X, \mathrm{E}[Y] = \mu_Y, \mathrm{Var}[X] = \sigma_X^2, \mathrm{Var}[Y] = \sigma_Y^2$
- Assume $0 < \sigma_X^2 < \infty$ and $0 < \sigma_Y^2 < \infty$

Definition 1.7 (Correlation coefficient).

The correlation of X and Y is the number defined by

$$\rho_{XY} = \frac{\mathrm{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

The value ρ_{XY} is also called the correlation coefficient.

Theorem 1.8 (Bounds of $\rho_{X,Y}$).

For any random variables X and Y,

(a) $-1 \leq \rho_{XY} \leq 1$

(b) $|\rho_{XY}| = 1$ *if and only if there exists numbers* $a \neq 0$ *and* b *such that* $P(Y = aX + b) = 1$*. If* $\rho_{XY} = 1$ *then* $a > 0$*, and if* $\rho_{XY} = -1$ *then* $a < 0$

Now we are ready to prove Theorem 1.5 (1/3)

Proof.

- Let $D_i = \frac{\partial}{\partial\theta} \log f(X_i \mid \theta) = \frac{\frac{\partial}{\partial\theta} f(X_i|\theta)}{f(X_i|\theta)}$ so that

$$D = \frac{\partial}{\partial\theta} \left\{ \log \prod_{i=1}^{n} f(X_i \mid \theta) \right\} = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i \mid \theta) = \sum_{i=1}^{n} D_i$$

Now we are ready to prove Theorem 1.5 (1/3)

Proof.

- Let $D_i = \frac{\partial}{\partial\theta}\log f(X_i \mid \theta) = \frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)}$ so that

$$D = \frac{\partial}{\partial\theta}\left\{\log\prod_{i=1}^{n} f(X_i \mid \theta)\right\} = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i \mid \theta) = \sum_{i=1}^{n} D_i$$

Since Theorem 1.8 implies $\{\mathrm{Cov}[W(\mathbf{X}), D]\}^2 \le \mathrm{Var}[W(\mathbf{X})]\,\mathrm{Var}[D]$ it follows that

$$\mathrm{Var}[W(\mathbf{X})] \ge \frac{\{\mathrm{Cov}[W(\mathbf{X}), D]\}^2}{\mathrm{Var}[D]}$$

Now we are ready to prove Theorem 1.5 (1/3)

Proof.

- Let $D_i = \frac{\partial}{\partial\theta}\log f(X_i \mid \theta) = \frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)}$ so that

$$D = \frac{\partial}{\partial\theta}\left\{\log\prod_{i=1}^{n} f(X_i \mid \theta)\right\} = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i \mid \theta) = \sum_{i=1}^{n} D_i$$

  Since Theorem 1.8 implies $\{\mathrm{Cov}[W(\mathbf{X}), D]\}^2 \le \mathrm{Var}[W(\mathbf{X})]\,\mathrm{Var}[D]$ it follows that

$$\mathrm{Var}[W(\mathbf{X})] \ge \frac{\{\mathrm{Cov}[W(\mathbf{X}), D]\}^2}{\mathrm{Var}[D]}$$

- Since $E[D] = \sum_{i=1}^{n} E[D_i] \overset{Thm.\ 1.6}{=} 0$, we have

$$\mathrm{Cov}[W(\mathbf{X}), D] = \mathrm{E}[W(\mathbf{X})D]$$

Proof of Theorem 1.5 (2/3)

$$\mathrm{E}_\theta W(\mathbf{X}) = k(\theta) = \int \cdots \int W(\boldsymbol{x}) \prod_{i=1}^{n} f(x_i \mid \theta)\, dx_1 \cdots dx_n$$

Proof of Theorem 1.5 (2/3)

$$\mathrm{E}_\theta W(\mathbf{X}) = k(\theta) = \int \cdots \int W(\boldsymbol{x}) \prod_{i=1}^{n} f(x_i \mid \theta)\, dx_1 \cdots dx_n$$

Differentiating with respect to θ, we obtain

$$\begin{aligned}
k'(\theta) &= \int \cdots \int W(\boldsymbol{x}) \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i \mid \theta)\, dx_1 \cdots dx_n \\
&= \int \cdots \int W(\boldsymbol{x}) \sum_{i=1}^{n} \left\{ \frac{\partial}{\partial\theta} f(x_i \mid \theta) \prod_{j \neq i} f(x_j \mid \theta) \right\} dx_1 \cdots dx_n \\
&= \int \cdots \int W(\boldsymbol{x}) \underbrace{\sum_{i=1}^{n} \left\{ \frac{\frac{\partial}{\partial\theta} f(x_i \mid \theta)}{f(x_i \mid \theta)} \right\}}_{D} f(\boldsymbol{x} \mid \theta)\, dx_1 \cdots dx_n \\
&= E[W(\mathbf{X})D]
\end{aligned}$$

Proof of Theorem 1.5 (3/3)

Furthermore, we have

$$
\begin{aligned}
\operatorname{Var}[D] &= \mathrm{E}\left[D^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^n D_i\right)^2\right] \\
&= \mathrm{E}\left[\sum_i D_i \sum_j D_j\right] = \mathrm{E}\left[\sum_i \sum_j D_i D_j\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}\left[D_i D_j\right] \\
&= \sum_{i=1}^n \mathrm{E}\left[D_i^2\right] + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{E}\left[D_i D_j\right] \\
&= \sum_{i=1}^n \mathrm{E}\left[\left(\frac{\partial}{\partial \theta} \log f(x \mid \theta)\right)^2\right] + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{E}\left[D_i D_j\right] \\
&= \sum_{i=1}^n I(\theta) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{E}\left[D_i\right] E\left[D_j\right] = nI(\theta) + 0.
\end{aligned}
$$

Proof of Theorem 1.5 (3/3)

Furthermore, we have

$$\begin{aligned}
\mathrm{Var}[D] &= \mathrm{E}\left[D^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^n D_i\right)^2\right] \\
&= \mathrm{E}\left[\sum_i D_i \sum_j D_j\right] = \mathrm{E}\left[\sum_i \sum_j D_i D_j\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}\left[D_i D_j\right] \\
&= \sum_{i=1}^n \mathrm{E}\left[D_i^2\right] + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{E}\left[D_i D_j\right] \\
&= \sum_{i=1}^n \mathrm{E}\left[\left(\frac{\partial}{\partial \theta} \log f(x \mid \theta)\right)^2\right] + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{E}\left[D_i D_j\right] \\
&= \sum_{i=1}^n I(\theta) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{E}\left[D_i\right] E\left[D_j\right] = nI(\theta) + 0.
\end{aligned}$$

Putting this all together, we have

$$\mathrm{Var}[W(\mathbf{X})] \geq \frac{\{\mathrm{Cov}[W(\mathbf{X}), D]\}^2}{\mathrm{Var}[D]} = \frac{\{k'(\theta)\}^2}{nI(\theta)}$$

. $\square$

A useful corollary

Corollary 1.9.

Under the assumptions of Theorem 1.5, if $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ is an unbiased estimator of θ, so that $k(\theta) = \theta$, then the Rao-Cramér inequality becomes

$$Var(W(\boldsymbol{X})) \geq \frac{1}{nI(\theta)}.$$

Poisson example revisited (again) I

Example 2.

Let $X_1, \ldots, X_n$ be iid Poisson (λ). Find the Cramér-Rao lower bound on the variance of unbiased estimators of λ. Also, find the MLE and show that it attains the Cramér-Rao lower bound. Since $\frac{\partial^2}{\partial \lambda^2} \log f(x \mid \lambda) =$

$$\frac{\partial^2}{\partial \lambda^2}\left[\log\left\{\lambda^x e^{-\lambda}(x!)^{-1}\right\}\right] = \frac{\partial^2}{\partial \lambda^2}[x \log \lambda - \lambda - \log(x!)] = -\frac{x}{\lambda^2}$$

we have

$$\mathrm{E}\left[\frac{\partial^2}{\partial \lambda^2} \log f(X \mid \lambda)\right] = \mathrm{E}\left[-\frac{1}{\lambda^2} X\right] = -\frac{1}{\lambda^2}\mathrm{E}[X] = -\frac{1}{\lambda^2}\lambda = -\frac{1}{\lambda}$$

By Theorem 1.6,

$$\mathrm{E}\left[\left(\frac{\partial}{\partial \theta} \log f(X \mid \theta)\right)^2\right] = -\mathrm{E}\left[\frac{\partial^2}{\partial \lambda^2} \log f(X \mid \lambda)\right] = \frac{1}{\lambda}$$

Poisson example revisited (again) II

Example 2.

So the Cramér-Rao lower bound for an unbiased estimator in the iid case is

$$\frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta[W(\boldsymbol{X})]\right)^2}{n\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right]} = \frac{1}{n\left(\frac{1}{\lambda}\right)} = \frac{\lambda}{n}$$

The MLE of λ *is* $\hat{\lambda} = \bar{X}$ *and* $\mathrm{Var}[\bar{X}] = \frac{\mathrm{Var}[X_1]}{n} = \frac{\lambda}{n}$ *so it attains the CRLB.*

CRLB for Normal Distribution Variance Estimator I

Example 3.

Let $X_1, \ldots, X_n$ be iid Normal $\left(\mu, \sigma^2\right)$ random variables. Find the Cramér-Rao lower bound on unbiased estimators of σ^2. Does S^2 satisfy the CRLB?

$$\begin{aligned}
\frac{\partial^2}{\partial\left(\sigma^2\right)^2} \log f\left(x \mid \mu, \sigma^2\right) &= \frac{\partial^2}{\partial\left(\sigma^2\right)^2}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}(x-\mu)^2\right] \\
&= \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}
\end{aligned}$$

Theorem 1.6 implies that

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f\left(X \mid \mu, \sigma^2\right)\right)^2\right] &= -\mathrm{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log f\left(X \mid \mu, \sigma^2\right)\right] \\
&= -\mathrm{E}\left[\frac{1}{2\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right]
\end{aligned}$$

CRLB for Normal Distribution Variance Estimator II

Example 3.

$$
\begin{aligned}
&= -\mathrm{E}\left[\frac{1}{2\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right] \\
&= -\frac{1}{2\sigma^4} + \frac{\mathrm{E}\left[(X-\mu)^2\right]}{\sigma^6} \\
&= -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} = \frac{1}{2\sigma^4}
\end{aligned}
$$

Thus, the CRLB is

$$\frac{1}{n\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right]} = \frac{2\sigma^4}{n}.$$

So, S^2 does not satisfy the CRLB since

$$\mathrm{Var}\left[S^2\right] = \frac{2\sigma^4}{n-1} = \frac{n}{n-1}\left(\frac{2\sigma^4}{n}\right) > \frac{2\sigma^4}{n} = CRLB$$

- In the previous example we are left with an incomplete answer; that is, is there a better unbiased estimator of σ^2 than S^2, or is the CRLB unattainable?

- In the previous example we are left with an incomplete answer; that is, is there a better unbiased estimator of σ^2 than S^2, or is the CRLB unattainable?
- The conditions for attainment of the CRLB are actually quite simple.
- Recall that the bound follows from an application of the Cauchy-Schwarz Inequality, so conditions for attainment of the bound are the conditions for equality in the Cauchy-Schwarz Inequality.
- The following Corollary is a useful tool because it gives us a way of finding a best unbiased estimator

- In the previous example we are left with an incomplete answer; that is, is there a better unbiased estimator of σ^2 than S^2, or is the CRLB unattainable?
- The conditions for attainment of the CRLB are actually quite simple.
- Recall that the bound follows from an application of the Cauchy-Schwarz Inequality, so conditions for attainment of the bound are the conditions for equality in the Cauchy-Schwarz Inequality.
- The following Corollary is a useful tool because it gives us a way of finding a best unbiased estimator

Corollary 1.10 (Attainment).

Let $X_1, \cdots, X_n$ be iid with pdf /pmf$f_X(x \mid \theta)$, where $f_X(x \mid \theta)$ satisfies the assumptions of the Cramer-Rao Theorem. Let $L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f_X(x_i \mid \theta)$ denote the likelihood function. If $W(\mathbf{X})$ is unbiased for $\tau(\theta)$, then $W(\mathbf{X})$ attains the Cramer-Rao lower bound if and only if

$$\frac{\partial}{\partial\theta} \log L(\theta \mid \mathbf{x}) = S_n(\mathbf{x} \mid \theta) = a(\theta)[W(\mathbf{X}) - \tau(\theta)]$$

for some function $a(\theta)$.

Continuation of Example 3

Is CRLB for σ^2 attainable?

$$
\begin{aligned}
L\left(\sigma^2 \mid \mathbf{x}\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(x_i-\mu\right)^2}{2\sigma^2}\right] \\
\log L\left(\sigma^2 \mid \mathbf{x}\right) &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \sum_{i=1}^{n} \frac{\left(x_i-\mu\right)^2}{2\sigma^2} \\
\frac{\partial \log L\left(\sigma^2 \mid \mathbf{x}\right)}{\partial \sigma^2} &= -\frac{n}{2}\frac{2\pi}{2\pi\sigma^2} + \sum_{i=1}^{n} \frac{\left(x_i-\mu\right)^2}{2\left(\sigma^2\right)^2} \\
&= -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{\left(x_i-\mu\right)^2}{2\sigma^4} \\
&= \frac{n}{2\sigma^4}\left(\frac{\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{n} - \sigma^2\right) \\
&= a\left(\sigma^2\right)\left(W(\mathbf{x}) - \sigma^2\right)
\end{aligned}
$$

Continuation of Example 3

Therefore,

1. If μ is known, the best unbiased estimator for σ^2 is $\sum_{i=1}^{n} (x_i - \mu)^2 / n$, and it attains the Cramer-Rao lower bound, i.e.

$$\mathrm{Var}\left[\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}\right] = \frac{2\sigma^4}{n}$$

2. If μ is not known, the Cramer-Rao lower-bound cannot be attained.

Bernoulli example I

Bernoulli example II

Example 4 (Bernoulli).

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim}$ Bernoulli (p). Is $\bar{X}$ the best unbiased estimator of p? Does it attain the Cramer-Rao lower bound?

$$
\begin{aligned}
L(p \mid \mathbf{x}) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
\log L(p \mid \mathbf{x}) &= \log \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
&= \sum_{i=1}^{n} \log \left[p^{x_i}(1-p)^{1-x_i} \right] \\
&= \sum_{i=1}^{n} \left[x_i \log p + (1-x_i)\log(1-p) \right] \\
&= \log p \sum_{i=1}^{n} x_i + \log(1-p)\left(n - \sum_{i=1}^{n} x_i \right)
\end{aligned}
$$

Bernoulli example III

Example 4 (Bernoulli).

$$
\begin{aligned}
\frac{\partial}{\partial p} \log L(p \mid \mathbf{x}) &= \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} \\
&= \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} \\
&= \frac{(1-p)n\bar{x} - np(1-\bar{x})}{p(1-p)} \\
&= \frac{n(\bar{x}-p)}{p(1-p)} \\
&= a(p)[W(\mathbf{x}) - \tau(p)]
\end{aligned}
$$

where $a(p) = \dfrac{n}{p(1-p)}$, $W(\mathbf{x}) = \bar{x}, \tau(p) = p$. *Therefore,* $\bar{X}$ *is the best unbiased estimator for* p *and attains the Cramer-Rao lower bound.*

Recap
Point Estimation
Maximum Likelihood Estimator
MLE example

Methods of Evaluating Estimators
Uniformly Minimum Variance Unbiased Estimator
Challenges in finding the UMVUE
Cramér–Rao Lower Bound (CRLB)
Fisher Information
Correlation review
CRLB proof
Attainment of CRLB

Rao-Blackwell

Methods for finding best unbiased estimator

- In the previous section, the concept of sufficiency was not used in our search for unbiased estimates. We will now see that consideration of sufficiency is a powerful tool, indeed.

Methods for finding best unbiased estimator

- In the previous section, the concept of sufficiency was not used in our search for unbiased estimates. We will now see that consideration of sufficiency is a powerful tool, indeed.
- The main theorem of this section, which relates sufficient statistics to unbiased estimates, is, as in the case of the Cramér-Rao Theorem, another clever application of some well-known theorems:
- Let X and Y be two random variables.
  - $E(X) = E[E(X \mid Y)]$
  - $\mathrm{Var}(X) = E[\mathrm{Var}(X \mid Y)] + \mathrm{Var}[E(X \mid Y)]$
  - $E[g(X) \mid Y] = \int_{x \in \mathcal{X}} g(x) f(x \mid Y)\, dx$ is a function of Y.
  - If X and Y are independent, $E[g(X) \mid Y] = E[g(X)]$.

Searching for a better unbiased estimator

Theorem 1.11 (Rao-Blackwell Theorem).

- *Suppose* $W(\mathbf{X})$ *is an unbiased estimator of* $\tau(\theta)$. *That is,* $E[W(\mathbf{X})] = \tau(\theta)$.

Searching for a better unbiased estimator

Theorem 1.11 (Rao-Blackwell Theorem).

- *Suppose* $W(\mathbf{X})$ *is an unbiased estimator of* $\tau(\theta)$*. That is,* $E[W(\mathbf{X})] = \tau(\theta)$*.*
- *Suppose* $T(\mathbf{X})$ *is any function of* $\mathbf{X} = (X_1, \cdots, X_n)$ *and is a sufficient statistic for* θ*.*

Searching for a better unbiased estimator

Theorem 1.11 (Rao-Blackwell Theorem).

- *Suppose* $W(\mathbf{X})$ *is an unbiased estimator of* $\tau(\theta)$*. That is,* $E[W(\mathbf{X})] = \tau(\theta)$.
- *Suppose* $T(\mathbf{X})$ *is any function of* $\mathbf{X} = (X_1, \cdots, X_n)$ *and is a sufficient statistic for* θ.
- *Define the estimator* $\phi(T) = E(W(\mathbf{X}) \mid T)$*. Then the following holds:*
  1. $E[\phi(T) \mid \theta] = \tau(\theta)$
  2. $\mathrm{Var}[\phi(T) \mid \theta] \leq \mathrm{Var}(W \mid \theta)$ *for all* θ.

  That is, $\phi(T)$ *is a uniformly better unbiased estimator of* $\tau(\theta)$.

Proof of Rao-Blackwell (Theorem 1.11)

1. $\tau(\theta) = E[W(\mathbf{X})] = E[E(W(\mathbf{X}) \mid T)] = E[\phi(T)] \quad (\text{ unbiased for } \tau(\theta))$

Proof of Rao-Blackwell (Theorem 1.11)

1. $\tau(\theta) = E[W(\mathbf{X})] = E[E(W(\mathbf{X}) \mid T)] = E[\phi(T)] \quad (\text{ unbiased for } \tau(\theta))$
2. $$\begin{aligned} \mathrm{Var}(W) &= \mathrm{Var}[E(W \mid T)] + E[\mathrm{Var}(W \mid T)] \\ &= \mathrm{Var}(\phi(T)) + E[\mathrm{Var}(W \mid T)] \\ &\geq \mathrm{Var}(\phi(T)) \quad (\text{ smaller variance than } W) \end{aligned}$$

Proof of Rao-Blackwell (Theorem 1.11)

1. $\tau(\theta) = E[W(\mathbf{X})] = E[E(W(\mathbf{X}) \mid T)] = E[\phi(T)]$ (unbiased for $\tau(\theta)$)
2. $$\begin{aligned} \mathrm{Var}(W) &= \mathrm{Var}[E(W \mid T)] + E[\mathrm{Var}(W \mid T)] \\ &= \mathrm{Var}(\phi(T)) + E[\mathrm{Var}(W \mid T)] \\ &\geq \mathrm{Var}(\phi(T)) \quad (\text{ smaller variance than } W) \end{aligned}$$
3. Need to show $\phi(T)$ is indeed an estimator.
$$\begin{aligned} \phi(T) &= E[W(\mathbf{X}) \mid T] \\ &= \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) f(\mathbf{x} \mid T) d\mathbf{x} \end{aligned}$$
Because T is a sufficient statistic, $f(\mathbf{x} \mid T)$ does not depend on θ. Therefore, $\phi(T) = \int_{\mathbf{x} \in \mathcal{X}} W(\mathbf{x}) f(\mathbf{x} \mid T) d\mathbf{x}$ does not depend on θ, and $\phi(T)$ is indeed an estimator of θ.

$\square$

Sufficiency in Action: Rao-Blackwellization

- The concept of data reduction without losing any information is an appealing idea. But one might ask what can sufficiency do for me apart from dimension reduction?

Sufficiency in Action: Rao-Blackwellization

- The concept of data reduction without losing any information is an appealing idea. But one might ask what can sufficiency do for me apart from dimension reduction?
- Given two decision rules (procedures) δ_1, δ_2, if one of them always had a smaller risk, we would naturally prefer that procedure.

Sufficiency in Action: Rao-Blackwellization

- The concept of data reduction without losing any information is an appealing idea. But one might ask what can sufficiency do for me apart from dimension reduction?
- Given two decision rules (procedures) δ_1, δ_2, if one of them always had a smaller risk, we would naturally prefer that procedure.
- The Rao-Blackwell theorem, proved independently by C.R. Rao and David Blackwell (Rao (1945), Blackwell (1947)), provides a concrete benefit of looking at sufficient statistics from a viewpoint of preferring procedures with lower risk.
- The Rao-Blackwell theorem says that after you have chosen your model, there is no reason to look beyond a minimal sufficient statistic

Rao-Blackwellization in Poisson

- Suppose based on $X_1, \cdots, X_n \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$, we wish to estimate $P_\lambda(X = 0) = e^{-\lambda}$.
- A layman's estimate might be the fraction of data values equal to zero:

$$\delta_1 \left(X_1, \cdots, X_n\right) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i = 0}$$

Rao-Blackwellization in Poisson

- Suppose based on $X_1, \cdots, X_n \stackrel{\text{iid}}{\sim} \text{Poi}(\lambda)$, we wish to estimate $P_\lambda(X = 0) = e^{-\lambda}$.
- A layman's estimate might be the fraction of data values equal to zero:

$$\delta_1 (X_1, \cdots, X_n) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i = 0}$$

- Rao-Blackwell theorem tells us we can do better by conditioning on $\sum_{i=1}^n X_i$, because $\sum_{i=1}^n X_i$ is sufficient in the iid Poisson case.

Rao-Blackwellization in Poisson

- To calculate this conditional expectation, recall that if $X_i, 1 \leq i \leq n$ are iid $\text{Poi}(\lambda)$, then any X_i given that $\sum_{i=1}^{n} X_i = t$ is distributed as $\text{Bin}\left(t, \frac{1}{n}\right)$

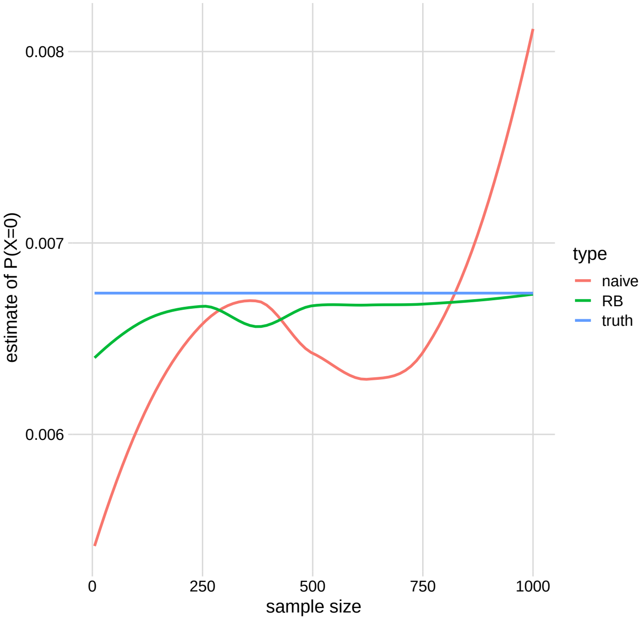
Rao-Blackwellization in Poisson

- To calculate this conditional expectation, recall that if $X_i, 1 \leq i \leq n$ are iid Poi(λ), then any X_i given that $\sum_{i=1}^n X_i = t$ is distributed as Bin $\left(t, \frac{1}{n}\right)$
- Then,

$$
\begin{aligned}
E_\lambda\left[\delta_1\left(X_1, \cdots, X_n\right) \mid \sum_{i=1}^n X_i = t\right] &= \frac{1}{n}\sum_{i=1}^n E_\lambda\left[I_{X_i=0} \mid \sum_{i=1}^n X_i = t\right] \\
&= \frac{1}{n}\sum_{i=1}^n P_\lambda\left[X_i = 0 \mid \sum_{i=1}^n X_i = t\right] \\
&= \frac{1}{n}\sum_{i=1}^n \left(1 - \frac{1}{n}\right)^t = \left(1 - \frac{1}{n}\right)^t
\end{aligned}
$$

- Thus, in the iid Poisson case, for estimating the probability of the zero value (no events), $\left(1 - \frac{1}{n}\right)^{\sum_{i=1}^n X_i}$ is better than the layman's estimator $\frac{1}{n}\sum_{i=1}^n I_{X_i=0}$

Simulation



True probability = P(X=0) where X follows a Poisson(5) distribution

Simulation

```
set.seed(334)
tt <- lapply(seq(5, 1000, by = 5), function(i){
n <- i
x <- rpois(n, lambda = 5)
data.frame(
        n = n,
        naive = mean(x == 0),
        RB = (1-1/n)^{sum(x)},
        truth = dpois(0,5))
})

ttt <- do.call(rbind,tt)

library(tidyverse)
library(cowplot)

ttt %>%
pivot_longer(cols = -1, names_to = "type") %>%
ggplot(mapping = aes(x = n, y = value, color = type)) +
  geom_smooth(se=FALSE) +
  cowplot::theme_minimal_grid() +
  xlab("sample size") +
  ylab("estimate of P(X=0)")+
  labs(caption = "True probability = P(X=0) where X follows a Poisson(5) distribution")
```