

More Maximum Likelihood Estimation!

Special Cases in this video:

- The Invariance Property of MLEs!!!

Suppose you have a random sample from a distribution with mean μ and variance σ^2 .

- An unbiased estimator of μ is \bar{X} .
- If we wanted to estimate μ^2 , Should we use \bar{X}^2 ?

Not if we want an unbiased estimator!

$$\begin{aligned} E[\bar{X}^2] &= \text{Var}[\bar{X}] + (E[\bar{X}])^2 \\ &= \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

The Invariance Property of MLEs

$$X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\text{rate} = \lambda)$$

Find the MLE of λ^2 .

Let $\tau = \lambda^2$.

Reparameterize the pdf:

$$f(x; \lambda) = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$$

Becomes:

$$f_2(x; \tau) = \sqrt{\tau} e^{-\sqrt{\tau} x} I_{(0, \infty)}(x)$$

The problem is now:

Let X_1, X_2, \dots, X_n be iid

from a distribution with pdf:

$$f_2(x; \tau) = \sqrt{\tau} e^{-\sqrt{\tau}x} \mathbf{I}_{(0, \infty)}(x)$$

Now this is a “straightforward” MLE problem!

You will get

$$\hat{\tau} = \frac{1}{\bar{X}^2}$$

The Invariance Property of MLEs

$$X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\text{rate} = \lambda)$$

Find the MLE of λ^2 .

Let $\tau = \lambda^2$.

The MLE for λ is $\hat{\lambda} = \frac{1}{\bar{X}}$

The MLE for τ is $\hat{\tau} = \frac{1}{\bar{X}^2} = \hat{\lambda}^2$

This is not a coincidence!

The Invariance Property of MLEs

$$f(x; \lambda) = \lambda e^{-\lambda x} \mathbf{I}_{(0, \infty)}(x)$$

$$f(\vec{x}; \lambda) \stackrel{\text{iid}}{=} \prod_{i=1}^n f(x_i, \lambda)$$

$$= \prod_{i=1}^n \lambda e^{-\lambda x_i} \mathbf{I}_{(0, \infty)}(x_i)$$

$$= \boxed{\lambda^n e^{-\lambda \sum_{i=1}^n x_i}} \prod_{i=1}^n \mathbf{I}_{(0, \infty)}(x_i)$$

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

If $\tau = \tau(\lambda) = \lambda^2$, we can rewrite this

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = \underbrace{(\sqrt{\tau})^n e^{-\sqrt{\tau} \sum_{i=1}^n x_i}}_{L(\tau)?}$$

$$L(\tau) = \tau^n e^{-\tau \sum_{i=1}^n x_i}$$

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

If $\tau = \tau(\lambda) = \lambda^2$, we can rewrite this

$$\begin{aligned} L(\lambda) &= \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = (\sqrt{\tau})^n e^{-\sqrt{\tau} \sum_{i=1}^n x_i} \\ &= \tilde{L}(\tau) \\ &= \tilde{L}(\tau(\lambda)) \end{aligned}$$

**Proof of invariance when τ is invertible
and L has a unique maximum:**

$$L(\boldsymbol{\theta}) = \tilde{L}(\tau(\boldsymbol{\theta}))$$

τ invertible $\Rightarrow \tau'(\boldsymbol{\theta}) > 0$ or $\tau'(\boldsymbol{\theta}) < 0$

$$\frac{d}{d\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \frac{d}{d\boldsymbol{\theta}} \tilde{L}(\tau(\boldsymbol{\theta})) = \frac{d}{d\tau} \tilde{L}(\tau(\boldsymbol{\theta})) \cdot \frac{d}{d\boldsymbol{\theta}} \tau(\boldsymbol{\theta})$$

$$\frac{d}{d\theta} L(\theta) = \frac{d}{d\tau} \tilde{L}(\tau(\theta)) \cdot \underbrace{\frac{d}{d\theta} \tau(\theta)}_{\substack{\text{never 0} \\ \text{since } \tau \text{ is} \\ \text{invertible}}}$$

$\frac{d}{d\theta} L(\theta)$ is zero when we plug in $\hat{\theta}$
by definition of the MLE

This implies that we must have

$$\frac{d}{d\tau} \tilde{L}(\tau(\hat{\theta})) = 0$$

$$\frac{d}{d\tau} \tilde{L}(\tau(\hat{\theta})) = 0$$

However,

$\frac{d}{d\tau} \tilde{L}(\tau(\theta))$ is zero when we plug in $\hat{\tau}(\theta)$ by definition of the MLE for $\tau(\theta)$

So

$$\frac{d}{d\tau} \tilde{L}(\tau(\hat{\theta})) = 0 = \frac{d}{d\tau} \tilde{L}(\hat{\tau}(\theta))$$

$$\frac{d}{d\tau} \tilde{L}(\tau(\hat{\theta})) = 0 = \frac{d}{d\tau} \tilde{L}(\hat{\tau}(\theta))$$

L has a unique maximum

$\Rightarrow \tilde{L}$ has a unique maximum

$$\Rightarrow \tau(\hat{\theta}) = \hat{\tau}(\theta)$$

The MLE of $\tau(\theta)$ is $\tau(\theta)$ with the MLE of θ plugged in!

Example:

$$X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$$

How can we estimate the probability that a typical measurement from this data set is greater than zero?

i.e. How can we estimate

$$p = P(X_i > 0) ?$$

One Answer:

$$\hat{p} = \frac{\# \text{ values in the sample that are } > 0}{n}$$

Example:

$$X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$$

How can we estimate the probability that a typical measurement from this data set is greater than zero?

Can we do this more formally with a maximum likelihood estimator?

$$\begin{aligned} P(X_i > 0) &= 1 - P(X_i = 0) \\ &= 1 - \frac{e^{-\lambda} \lambda^0}{0!} = \underbrace{1 - e^{-\lambda}}_{\tau(\lambda)} \end{aligned}$$

$$X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$$

The pdf is:

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \mathbf{I}_{\{0,1,2,\dots\}}(x)$$

The joint pdf is:

$$f(\vec{x}; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \mathbf{I}_{\{0,1,2,\dots\}}(x_i)$$

$$= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i!)} \prod_{i=1}^n \mathbf{I}_{\{0,1,2,\dots\}}(x_i)$$

$$X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$$

A likelihood is:

$$L(\lambda) = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}$$

The log-likelihood is:

$$\ell(\lambda) = -n\lambda + \left(\sum_i^n x_i \right) \ln \lambda$$

$$\frac{d}{d\lambda} \ell(\lambda) = 0 \quad \Rightarrow \quad \hat{\lambda} = \bar{X}$$

$$X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$$

The MLE for λ is $\hat{\lambda} = \bar{X}$

By the invariance property of MLEs,
the MLE for $p = \tau(\lambda) = 1 - e^{-\lambda}$ is

$$\hat{\tau}(\lambda) \overset{\text{invar}}{=} \tau(\hat{\lambda}) = 1 - e^{-\bar{X}}$$

$$\hat{p}_1 = \frac{\text{\# values in the sample that are } > 0}{n}$$

$$\hat{p}_2 = 1 - e^{-\bar{X}}$$

Which is better?

$$\hat{p}_1 = \frac{1}{n} \sum_{i=1}^n I_{\{X_i > 0\}}$$

$$E[\hat{p}_1] = \frac{1}{n} \sum_{i=1}^n E[I_{\{X_i > 0\}}] = E[I_{\{X_1 > 0\}}]$$

$$E[I_{\{X_1 > 0\}}] = 0 \cdot P(I_{\{X_i > 0\}} = 0) + 1 \cdot P(I_{\{X_i > 0\}} = 1)$$

$$= P(I_{\{X_i > 0\}} = 1)$$

$$= P(X_i > 0) = p = 1 - e^{-\lambda}$$

$$E[\hat{p}_1] = E[I_{\{X_1 > 0\}}] = p = 1 - e^{-\lambda}$$

\hat{p}_1 is an unbiased estimator of p .

$$E[\hat{p}_2] = E[1 - e^{-\bar{X}}]$$

$$= 1 - E[e^{-\bar{X}}] \quad ?$$

- **Method One:**

$$E[e^{-\bar{X}}] = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \dots \sum_{x_n=0}^{\infty} e^{-\frac{1}{n} \sum x_i} \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod (x_i!)}$$

No thanks...

$$E[\hat{p}_2] = 1 - E[e^{-\bar{X}}] = 1 - E[e^{-Y/n}]$$

- **Method Two:**

$$\text{Let } Y = \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$$

Use moment generating functions, we can show that $Y \sim \text{Poisson}(n\lambda)$

$$E[e^{-\bar{X}}] = E[e^{-\frac{1}{n}Y}] = \sum_{y=0}^{\infty} e^{-\frac{1}{n}y} \frac{e^{-n\lambda}(n\lambda)^y}{y!}$$

No thanks...

$$E[\hat{p}_2] = E[1 - e^{-\bar{X}}] = 1 - E[e^{-\bar{X}}]$$

- Method Three:

Find the distribution of \bar{X} .

Let $W = \bar{X}$.

$$\begin{aligned} P(W = w) &= P(\bar{X} = w) = P\left(\frac{1}{n}Y = w\right) \\ &= P(Y = nw) = \frac{e^{-n\lambda}(n\lambda)^{nw}}{(nw)!} \end{aligned}$$

$$E[e^{-\bar{X}}] = E[e^{-W}] = \sum e^w \cdot P(W = w)$$

No thanks... ? 0, 1/n, 2/n, 3/n, ...

$$E[\hat{p}_2] = E[1 - e^{-\bar{X}}] = 1 - E[e^{-\bar{X}}]$$

- **Method Four:**

Let $Y = \sum_{i=1}^n X_i$.

- We know that $Y \sim \text{Poisson}(n\lambda)$.

- So, we know the mgf for Y is

$$M_Y(t) = \exp[n\lambda(e^t - 1)]$$

Now

$$E[e^{-\bar{X}}] = E[e^{-\frac{1}{n}Y}] = M_Y\left(-\frac{1}{n}\right)$$

$$\begin{aligned} E[e^{-\bar{X}}] &= E\left[-\frac{1}{n}Y\right] = M_Y\left(-\frac{1}{n}\right) \\ &= \exp[n\lambda(e^{-1/n} - 1)] \end{aligned}$$

So

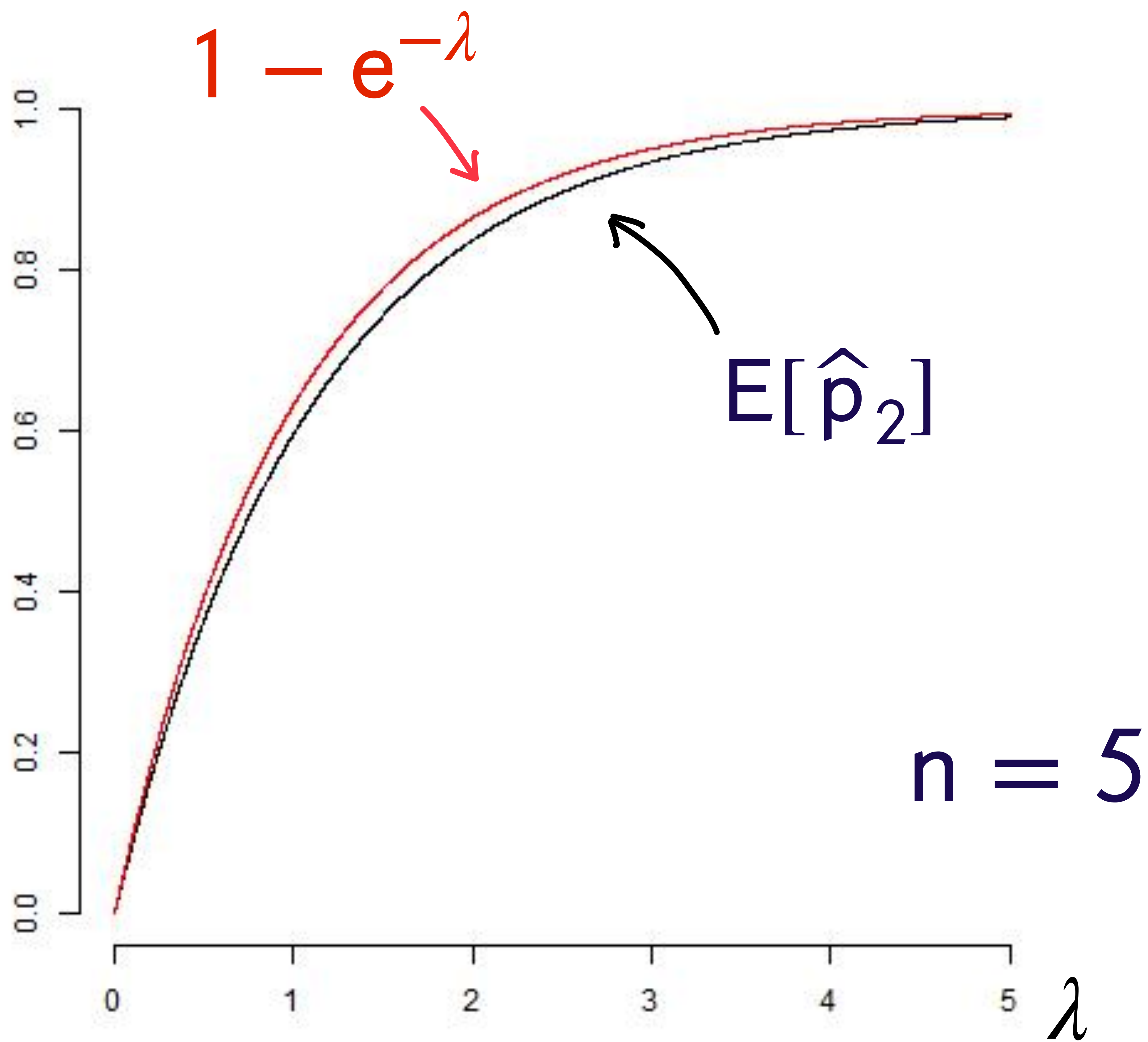
$$\begin{aligned} E[\hat{p}_2] &= E[1 - e^{-\bar{X}}] \\ &= 1 - \exp[n\lambda(e^{-1/n} - 1)] \end{aligned}$$

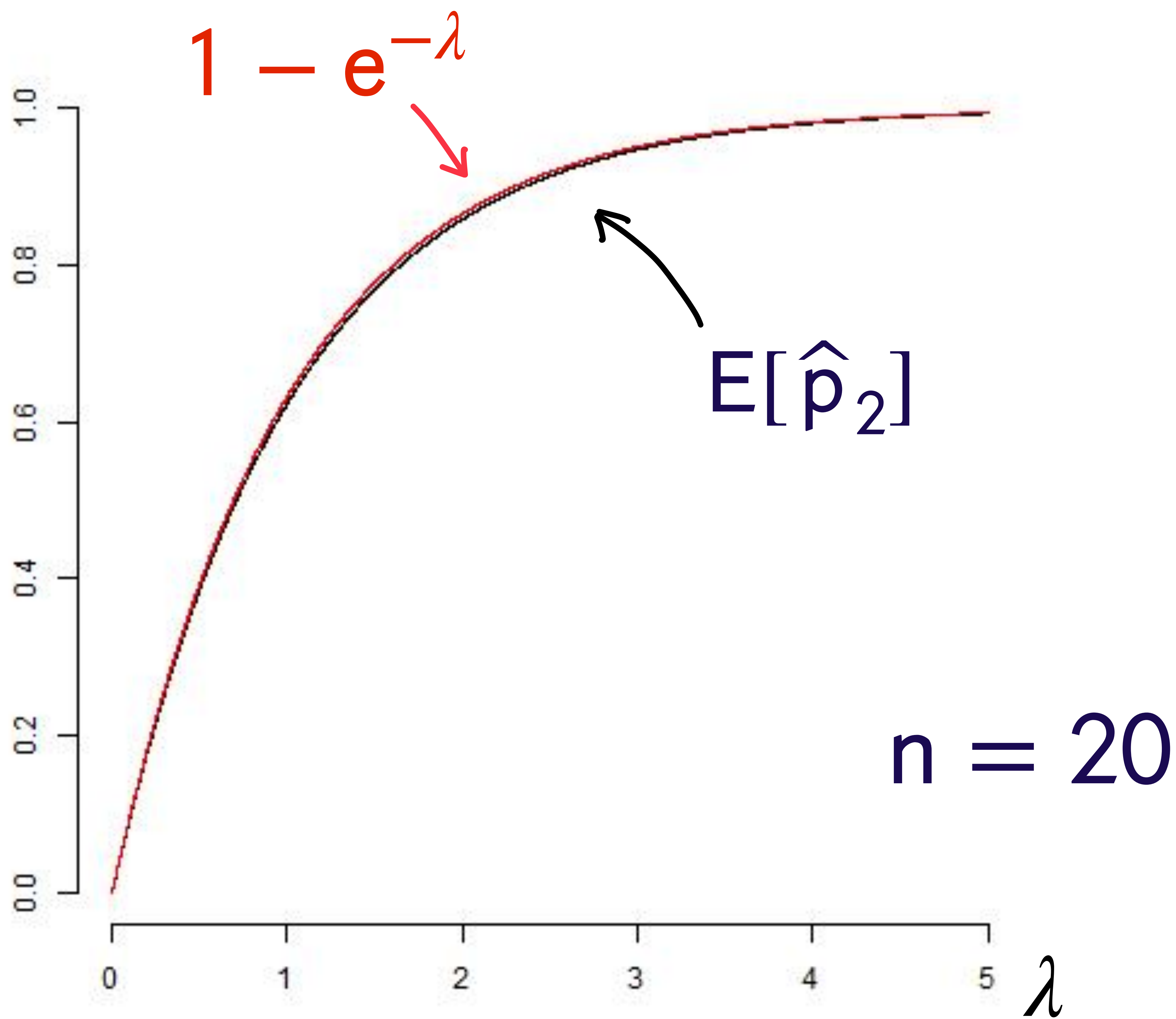
$$\begin{aligned} E[e^{-\bar{X}}] &= E\left[-\frac{1}{n}Y\right] = M_Y\left(-\frac{1}{n}\right) \\ &= \exp[n\lambda(e^{-1/n} - 1)] \end{aligned}$$

So

$$\begin{aligned} E[\hat{p}_2] &= E[1 - e^{-\bar{X}}] \\ &= 1 - \exp[n\lambda(e^{-1/n} - 1)] \end{aligned}$$

We want this to be $1 - e^{-\lambda}$.





In summary, our estimates of $p = 1 - e^{-\lambda}$

- $\hat{p}_1 = \frac{1}{n} \sum_{i=1}^n I_{\{X_i > 0\}}$

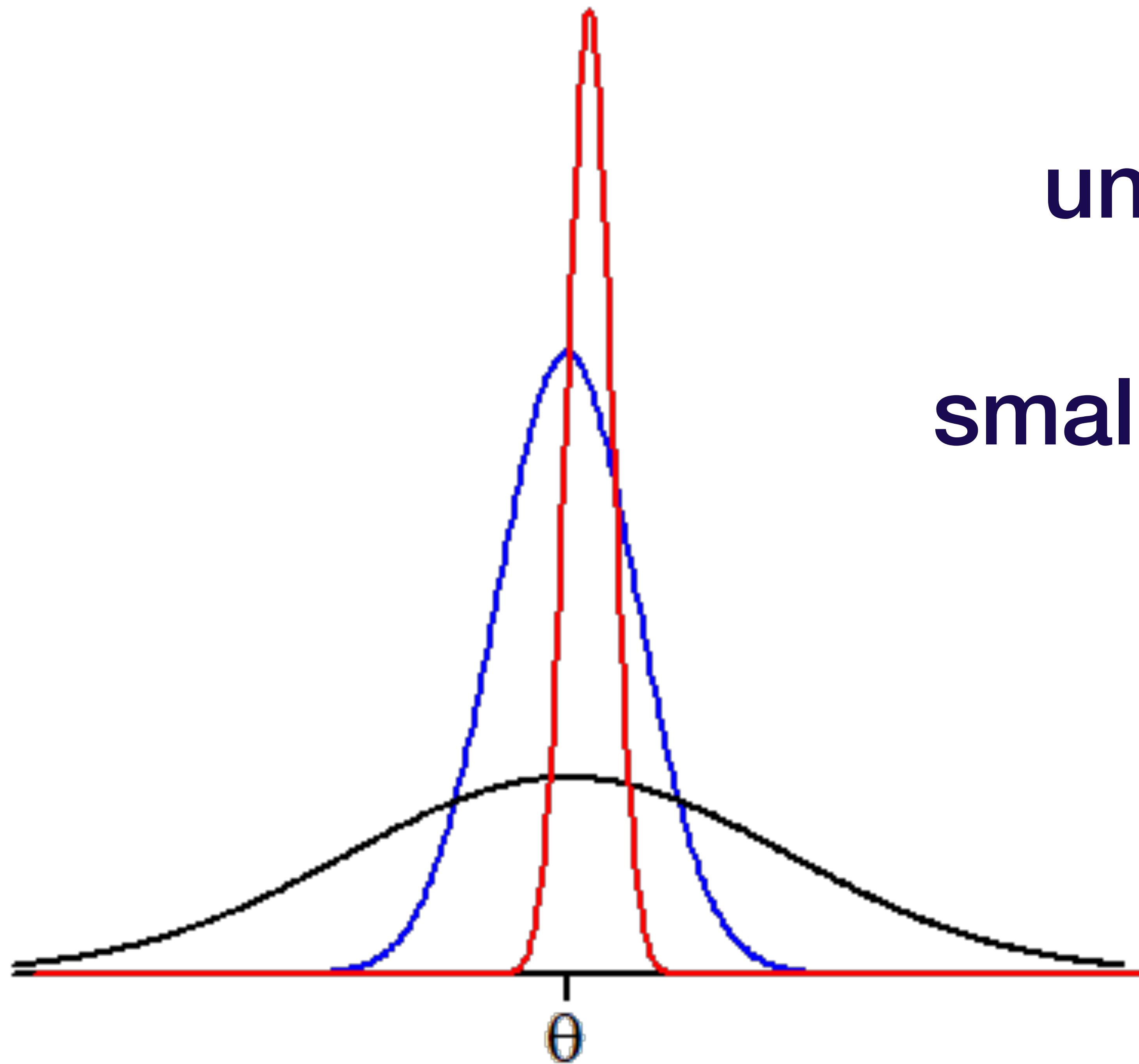
is an unbiased estimator of p

- $\hat{p}_2 = 1 - e^{-\bar{X}}$

is expected to be a little below p

(negative “bias”)

Should we compare the variances of \hat{p}_1
and \hat{p}_2 ?



unbiased
vs
small variance