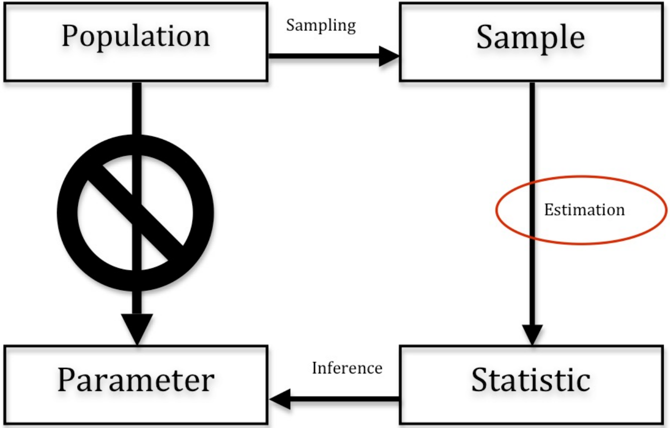
PLSC 502 – Fall 2023
Estimation and Estimators

October 16, 2023

Remember This?



Random Variables, Take Two

For a random variable X:

$$X_i = \underbrace{\mu}_{\text{"systematic part"}} + \underbrace{u_i}_{\text{"stochastic part"}}$$

where μ is the population mean (expected value) of X and $\text{Cov}(\mu, u) = 0$.

That implies that:

$$\underbrace{u_i}_{\text{"error"}} = \underbrace{X_i}_{\text{"observed"}} - \underbrace{\mu}_{\text{"expected"}}$$

Random Variables, Take Two

What's our expectation for u?

$$\begin{aligned}\mathrm{E}(u) &= \mathrm{E}(X-\mu)\\ &= \mathrm{E}(X)-\mathrm{E}(\mu)\\ &= \mathrm{E}(X)-\mu\\ &= \mu-\mu\\ &= 0\end{aligned}$$

and so:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2]\\ &= \mathrm{E}(u^2)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(u) &= \mathrm{E}[(u-E(u))^2]\\ &= \mathrm{E}[(u-0)^2]\\ &= \mathrm{E}(u^2).\end{aligned}$$

Estimation Example: $\bar{X}$

Challenge: Estimate $\mu = E(X)$ from a sample of N observations.

$$
\begin{aligned}
\bar{X} &= \frac{1}{N}\sum_{i=1}^{N} X_i \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mu + u_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mu) + \frac{1}{N}\sum_{i=1}^{N}(u_i) \\
&= \frac{1}{N}(N\mu) + \frac{1}{N}\sum_{i=1}^{N}(u_i) \\
&= \mu + \bar{u}
\end{aligned}
$$

The point: $\bar{X}$ *is a random variable*.

Small-Sample Properties

- Hold irrespective of N
- "Small sample estimators"

Large-Sample (Asymptotic) Properties

- Hold as $N \rightarrow \infty$
- "More is better"

Unbiasedness

Start with a generic population parameter θ, and an estimator of it $\hat{\theta}$ based on a sample of N observations...

<u>Unbiasedness</u> means:

$$\mathsf{E}(\hat{\theta}) = \theta$$

"Bias" is:

$$B(\hat{\theta}) = \mathsf{E}(\hat{\theta}) - \theta$$

Example: For $\bar{X}$, we know that:

$$\begin{aligned} \mathsf{E}(\bar{X}) &= \mathsf{E}(\mu + \bar{u}) \\ &= \mathsf{E}(\mu) + \mathsf{E}(\bar{u}) \\ &= \mu + 0 \\ &= \mu \end{aligned}$$

and so:

$$B(\bar{X}) = 0.$$

Multiple Unbiased Estimators

For $N = 2$:

$$Z = \lambda_1 X_1 + \lambda_2 X_2.$$

note that

$$\begin{aligned}
\mathsf{E}(Z) &= \mathsf{E}(\lambda_1 X_1 + \lambda_2 X_2) \\
&= \mathsf{E}(\lambda_1 X_1) + \mathsf{E}(\lambda_2 X_2) \\
&= \lambda_1 \mathsf{E}(X_1) + \lambda_2 \mathsf{E}(X_2) \\
&= \lambda_1 \mu + \lambda_2 \mu \\
&= (\lambda_1 + \lambda_2)\mu
\end{aligned}$$

Means

$$\mathsf{E}(Z) = \mu \;\longleftrightarrow\; (\lambda_1 + \lambda_2) = 1.0$$

and in fact:

$$\mathsf{E}(Z) = \mu \;\longleftrightarrow\; \sum_{i=1}^{N} \lambda_i = 1.0.$$

Q: Why do we use $\lambda_i = \frac{1}{N}\ \forall\ i$?

Efficiency:

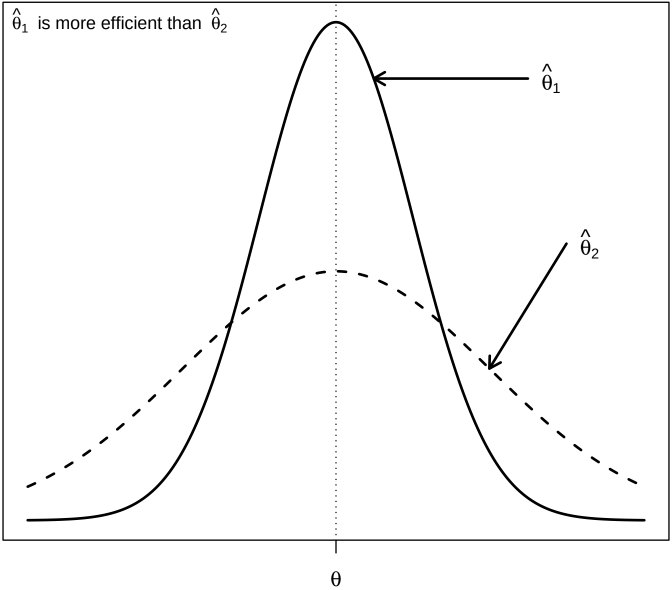
- is *relative variability* – how much difference we would expect in our $\hat{\theta}$s from one sample to the next...
- ...so a more efficient estimator has higher "reliability."
- ...is related to **information** (specifically, the *Fisher information* in the sample).

Note that:

- To be *fully efficient*[1], an estimator must be unbiased. BUT...
- ...the least-variance estimator need not be an unbiased one.

[1] That is, to achieve the *Cramer-Rao lower bound*, something we'll discuss in detail a bit later.

Efficiency: Unbiased $\hat{\theta}$s



Efficiency (continued)

Note that for our example with $N = 2$, where $\text{Var}(X) = \sigma^2$:

$$\begin{aligned}\text{Var}(Z) &= \text{Var}(\lambda_1 X_1 + \lambda_2 X_2) \\ &= (\lambda_1^2 + \lambda_2^2)\sigma^2\end{aligned}$$

and:

$$\begin{aligned}\lambda_1^2 + \lambda_2^2 &= \lambda_1^2 + (1 - \lambda_1)^2 \\ &= \lambda_1^2 + (1 - 2\lambda_1 + \lambda_1^2) \\ &= 2\lambda_1^2 - 2\lambda_1 + 1.\end{aligned}$$

Minimize!

$$\frac{\partial 2\lambda_1^2 - 2\lambda_1 + 1}{\partial \lambda_1} = 4\lambda_1 - 2$$

$$\begin{aligned}4\lambda_1 - 2 &= 0 \\ \lambda_1 &= 0.5\end{aligned}$$

Mean Squared Error

The "mean squared error" ("MSE") of an estimator $\hat{\theta}$ is:

$$\begin{aligned}\mathsf{MSE}(\hat{\theta}) &= \mathsf{E}[(\hat{\theta}-\theta)^2] \\ &= \mathsf{E}[B(\hat{\theta})^2] \\ &= \mathsf{Var}(\hat{\theta}) + [B(\hat{\theta})^2]\end{aligned}$$

Note that:

- The MSE of an unbiased estimator is equal to its variance [that is, $\mathsf{MSE} = \mathsf{Var}(\hat{\theta})$].
- Among unbiased estimators, the efficient estimator will always have the smallest MSE [because $B(\hat{\theta}) = [B(\hat{\theta})]^2 = 0$].

Comparing Estimators via MSE

As an estimator of μ, $\bar{X}$ has:

- $B(\bar{X}) = 0$
- $\mathrm{Var}(\bar{X}) = \sigma^2/N$, so
- $\mathrm{MSE}(\bar{X}) = \sigma^2/N + (0)^2 \; = \; \sigma^2/N$.

My alternative: the "Six Estimator"!

$$\zeta = 6$$

(That's a "zeta." Gotta learn your Greek letters.)

Comparing Estimators via MSE

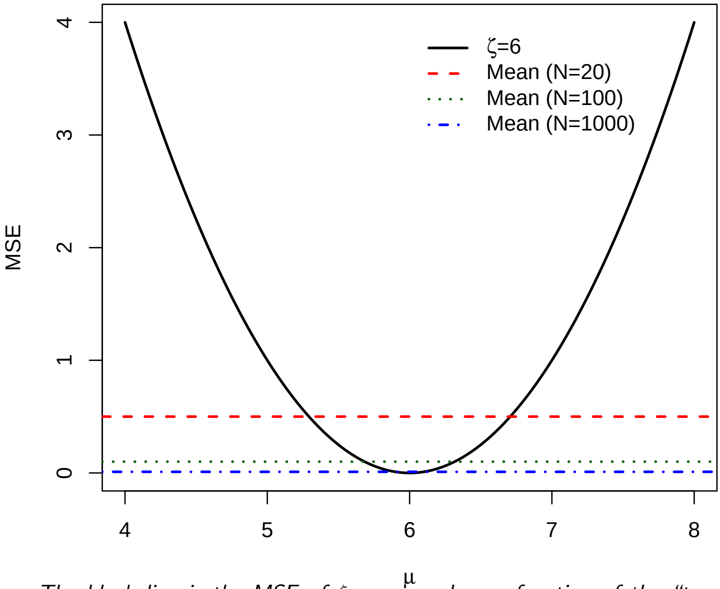
Properties of ζ (for $\zeta = 6$):

$$\begin{aligned} B(\zeta) &= \mathsf{E}(\zeta - \mu) \\ &= \mathsf{E}(6) - \mathsf{E}(\mu) \\ &= 6 - \mu, \end{aligned}$$

$$\begin{aligned} \mathsf{Var}(\zeta) &= \mathsf{Var}(6) \\ &= 0 \end{aligned}$$

and so:

$$\begin{aligned} \mathsf{MSE}(\zeta) &= \mathsf{Var}(\zeta) + [B(\zeta)]^2 \\ &= 0 + (6 - \mu)^2 \\ &= 36 - 12\mu + \mu^2 \end{aligned}$$



The black line is the MSE of ζ, expressed as a function of the "true" population mean μ. The other colored lines are the MSEs for $\bar{X}$, under the assumption that $\sigma^2 = 10$ and $N = \{20, 100, 1000\}$, respectively.

Large-Sample Properties: Consistency

An estimator $\hat{\theta}$ is *consistent* if:

$$\lim_{N \to \infty} \Pr[|\hat{\theta} - \theta| < \epsilon] = 1.0$$

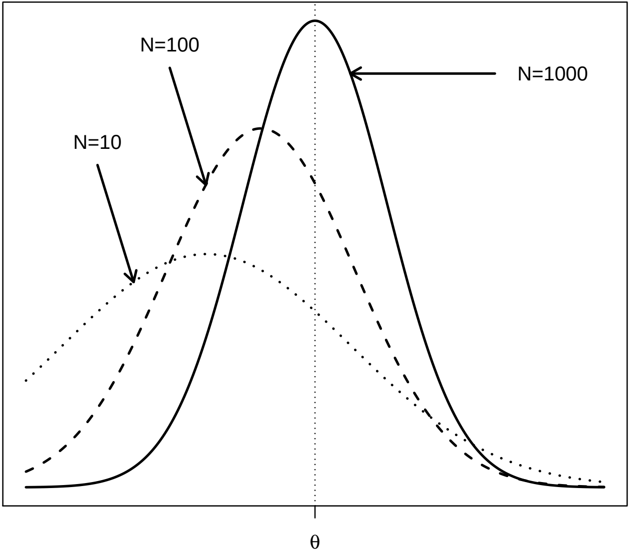
for an arbitrarily small $\epsilon > 0$

Equivalently:

$$\mathrm{E}(\hat{\theta}_N) \to \theta \text{ as } N \to \infty$$

Intuition: "Asymptotic unbiasedness"...

A Consistent Estimator $\hat{\theta}$



Estimation, Generally

Among estimators:

- Unbiased $>$ Consistent $>$ Biased
- Fully Efficient $>$ Asymptotically Efficient $>$ Inefficient
- MSE is <u>one</u> way to trade off bias vs. efficiency

Estimation Example: The Poisson

Recall the *Poisson* distribution:

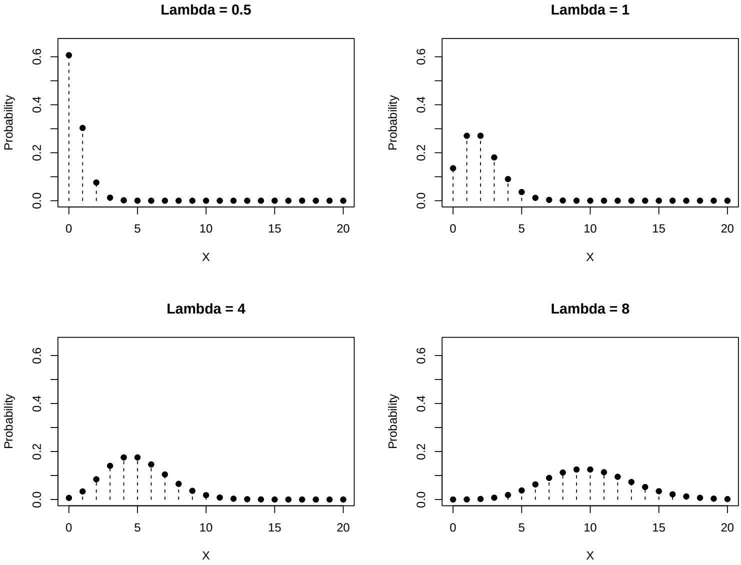
$$f(x) \equiv \Pr(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!}.$$

for $x \in \{0, 1, 2, ...\}$.

The Poisson:

- ...is a distribution for *counts* of *independent events*;
- ...is a *one parameter* distribution, where
- ...the parameter λ is both the *mean* and the *variance* of X.

Poisson Densities



Poisson Estimation

What is a "good" estimator for λ?

For a series of N i.i.d. values $\{X_1, X_2, ...X_N\}$ drawn from a Poisson distribution, their *joint* probability is:

$$f(X_1, X_2, ...X_N|\lambda) \equiv f(\mathbf{X}) = \prod_{i=1}^{N} \frac{\lambda^{X_i} \exp(-\lambda)}{X_i!}. \qquad (1)$$

This is sometimes known as the *likelihood* (more on that later...), and it relies on the fact that the joint probability of two independent random variables equals the product of the two marginal probabilities:

$$\Pr(A, B \mid A \perp B) = \Pr(A) \times \Pr(B)$$

Poisson Estimation

We can simplify (1) by taking its log:

$$\begin{aligned}
\ln[f(\mathbf{X})] &= \ln\left[\prod_{i=1}^{N}\frac{\lambda^{X_i}\exp(-\lambda)}{X_i!}\right] \\
&= \sum_{i=1}^{N}\ln\left[\frac{\lambda^{X_i}\exp(-\lambda)}{X_i!}\right] \\
&= \sum_{i=1}^{N}\left[X_i\ln(\lambda)-\lambda-\ln(X_i!)\right] \\
&= -N\lambda+\ln(\lambda)\sum_{i=1}^{N}X_i-\sum_{i=1}^{N}\ln(X_i!)
\end{aligned}$$

(This is the *log-likelihood*...)

Poisson Estimation

If we want to know the value of λ that maximizes this joint (log-)probability, we can figure that out too:

$$\frac{\partial \ln f(\mathbf{X})}{\partial \lambda} = -N + \frac{1}{\lambda} \sum_{i=1}^{N} X_i$$

and then:

$$-N + \frac{1}{\lambda} \sum_{i=1}^{N} X_i = 0$$

and so:

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

IOW, one version of a "good" estimator for λ (the "maximum likelihood estimator") is the empirical mean $\bar{X}$...

Poisson Mean Characteristics

What can we say about this $\hat{\lambda}$?

$$\begin{aligned} E(\hat{\lambda}) &= E\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} E(X_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} \lambda \\ &= \lambda \end{aligned}$$

so:

$$B(\hat{\lambda}) = 0 \text{ (unbiasedness)}$$

Also: Because $\mathrm{Var}(X) = \lambda$, this also means that $\hat{\lambda}$ is also an unbiased estimate of the variance.

More Poisson Mean Characteristics

Variance / efficiency?

Because $\hat{\lambda}$ is unbiased, we know that:

$$\mathsf{MSE}(\hat{\lambda}) = \mathsf{Var}(\hat{\lambda}).$$

Central limit theorem means that:

$$\hat{\lambda} \sim N\left(\lambda, \frac{\lambda}{N}\right)$$

so:

$$\mathsf{MSE}(\hat{\lambda}) = \frac{\lambda}{N}.$$

Example One: Simulation

The Plan:

1. Draw N values of X from a Poisson distribution with a known value of λ;
2. Calculate $\hat{\lambda} = \bar{X}$;
3. Repeat steps (1) - (2) many times;
4. Examine the distribution of the $\hat{\lambda}$s

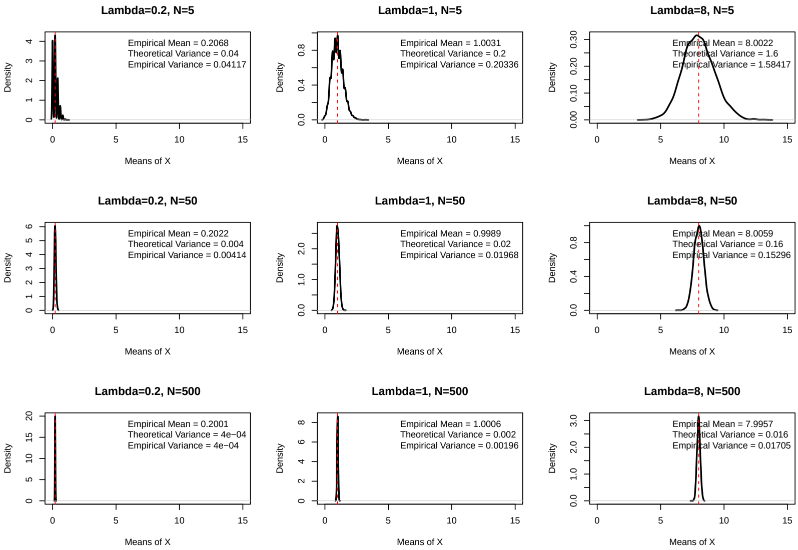
Details

- Vary $\lambda \ \in \ \{0.2, 1.0, 8.0\}$
- Vary $N \ \in \ \{5, 50, 500\}$

A Little Code

```
> L<-c(0.2,1,8)  # the lambdas:
> N<-c(5,50,500) # the Ns:
> sims<-4000     # number of sims
> Out<-data.frame(matrix(nrow=sims,ncol=length(N)*length(L)))
>
> c <- 0                        # column indicator for "Out"
> set.seed(7222009)             # Seed
>
> for(i in 1:length(N)) {      # Looping over sample sizes...
+   for(j in 1:length(L)) {  # Looping over lambdas
+     c <- c+1                  # increment column indicator
+     for(k in 1:sims) {       # Looping over 4000 simulations each
+       df<-rpois(N[i],L[j]) # Draw N values from Poisson(lambda)
+       Out[k,c]<-mean(df)   # Store the mean of the N draws
+       rm(df)
+     }
+   }
+ }
```

A Picture



Lambda=0.2, N=5
Empirical Mean = 0.2068
Theoretical Variance = 0.04
Empirical Variance = 0.04117
Lambda=1, N=5
Empirical Mean = 1.0031
Theoretical Variance = 0.2
Empirical Variance = 0.20336
Lambda=8, N=5
Empirical Mean = 8.0022
Theoretical Variance = 1.6
Empirical Variance = 1.58417
Lambda=0.2, N=50
Empirical Mean = 0.2022
Theoretical Variance = 0.004
Empirical Variance = 0.00414
Lambda=1, N=50
Empirical Mean = 0.9989
Theoretical Variance = 0.02
Empirical Variance = 0.01968
Lambda=8, N=50
Empirical Mean = 8.0059
Theoretical Variance = 0.16
Empirical Variance = 0.15296
Lambda=0.2, N=500
Empirical Mean = 0.2001
Theoretical Variance = 4e−04
Empirical Variance = 4e−04
Lambda=1, N=500
Empirical Mean = 1.0006
Theoretical Variance = 0.002
Empirical Variance = 0.00196
Lambda=8, N=500
Empirical Mean = 7.9957
Theoretical Variance = 0.016
Empirical Variance = 0.01705
Density
Means of X

Example Two: "Real" Data

Back to the English Premier League!

```
> PL

   Rank                        Team GamesPlayed Won Drew Lost GoalsFor GoalsAgainst GoalDifference Points
2     1           Tottenham Hotspur           8   6    2    0       18            8             10     20
3     2                     Arsenal           8   6    2    0       16            6             10     20
4     3             Manchester City           8   6    0    2       17            6             11     18
5     4                   Liverpool           8   5    2    1       18            9              9     17
6     5                 Aston Villa           8   5    1    2       19           12              7     16
7     6 Brighton and Hove Albion              8   5    1    2       21           16              5     16
8     7             West Ham United           8   4    2    2       15           12              3     14
9     8            Newcastle United           8   4    1    3       20            9             11     13
10    9              Crystal Palace           8   3    3    2        7            7              0     12
11   10           Manchester United           8   4    0    4        9           12             -3     12
12   11                     Chelsea           8   3    2    3       11            7              4     11
13   12                      Fulham           8   3    2    3        8           13             -5     11
14   13           Nottingham Forest           8   2    3    3        8           10             -2      9
15   14     Wolverhampton Wanderers           8   2    2    4        9           14             -5      8
16   15                   Brentford           8   1    4    3       11           12             -1      7
17   16                     Everton           8   2    1    5        9           12             -3      7
18   17                  Luton Town           8   1    1    6        6           15             -9      4
19   18                     Burnley           8   1    1    6        7           20            -13      4
20   19                 Bournemouth           8   0    3    5        5           18            -13      3
21   20            Sheffield United           8   0    1    7        6           22            -16      1
```

Premier League: Summary

```
> psych::describe(PL)

               vars  n  mean   sd median trimmed  mad min max range  skew kurtosis   se
Rank*             1 20 10.50 5.92   10.5   10.50 7.41   1  20    19  0.00    -1.38 1.32
Team*             2 20 10.50 5.92   10.5   10.50 7.41   1  20    19  0.00    -1.38 1.32
GamesPlayed       3 20  8.00 0.00    8.0    8.00 0.00   8   8     0   NaN      NaN 0.00
Won               4 20  3.15 1.98    3.0    3.19 2.97   0   6     6 -0.04    -1.37 0.44
Drew              5 20  1.70 1.03    2.0    1.69 1.48   0   4     4  0.31    -0.57 0.23
Lost              6 20  3.15 1.93    3.0    3.12 1.48   0   7     7  0.26    -0.81 0.43
GoalsFor          7 20 12.00 5.37   10.0   11.75 5.93   5  21    16  0.32    -1.58 1.20
GoalsAgainst      8 20 12.00 4.53   12.0   11.62 4.45   6  22    16  0.55    -0.62 1.01
GoalDifference    9 20  0.00 8.49   -0.5    0.44 9.64 -16  11    27 -0.29    -1.17 1.90
Points           10 20 11.15 5.71   11.5   11.19 6.67   1  20    19 -0.09    -1.22 1.28
```

Fitting a Poisson Distribution

```
> library(MASS)
> PoisMean
    lambda
  1.7000000
 (0.2915476)

> # Components:

> coef(PoisMean)
lambda
   1.7

> vcov(PoisMean)
       lambda
lambda  0.085

> # Note also:
>
> coef(PoisMean) / nrow(PL)
lambda
 0.085

> # and:
>
> (PoisMean$sd)^2
lambda
 0.085
```

Actual vs. Theoretical Draws