6.434J/16.391J
Statistics for Engineers and Scientists — February 28
MIT, Spring 2006 — Handout #5

Solution 2

Problem 1 Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,2)$ be independent random variables.

(a) What is a relationship between positive real numbers, a and b, such that $\mathbb{P}\left\{X^2 \leq a\right\} = \mathbb{P}\left\{Y^2 \leq b\right\}$?

(b) Use MATLAB to obtain a and b such that $\mathbb{P}\left\{X^2 \leq a\right\} = \mathbb{P}\left\{Y^2 \leq b\right\} = 0.95$. Explain briefly how you obtain a and b.
[*Hint*: This problem can be solved by a single MATLAB command. Use a function that inverses an appropriate distribution.]

Solution

(a) Define $Z \triangleq Y/\sqrt{2}$. Thus Z is a $\mathcal{N}(0,1)$ normal random variable, same as X. Then from the given relationship involving a and b,

$$\begin{aligned}\mathbb{P}\left\{X^2 \leq a\right\} &= \mathbb{P}\left\{Y^2 \leq b\right\}\\ &= \mathbb{P}\left\{2Z^2 \leq b\right\}\\ &= \mathbb{P}\left\{Z^2 \leq b/2\right\}\\ &= \mathbb{P}\left\{X^2 \leq b/2\right\}\end{aligned}$$

Therefore,

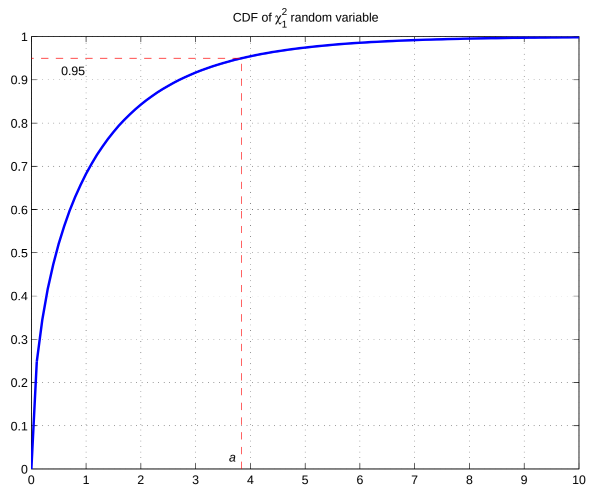
$$F_{X^2}(a) = F_{X^2}(b/2).$$

Note that X^2 is a chi-squared random variable with 1 degree of freedom, and F_{X^2} is its cdf. Since the cdf of a chi-square random variable (with any degree of freedom) is *strictly* increasing on $(0, \infty)$, the cdf is one-to-one. That is, $a = b/2$.

(b) A MATLAB command to obtain a is as follows:

```
% Return ‘a’ such that "Pr{ V <= a } = 0.95",
% where V is a chi-square r.v. with 1 degree of freedom.
a = chi2inv( 0.95, 1 );
```

The execution yields $a = 3.8415$, and thus $b = 2a = 7.6830$.



Problem 2 Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) $\mathcal{N}(\mu, \mu)$ random variables for some $\mu > 0$.

(a) Find an unbiased estimator of the mean μ.

(b) Find an unbiased estimator of the variance μ that is independent of the estimator in (a).

(c) Find an unbiased estimator of μ^2.
[*Hint*: Use results from (a) and (b).]

Solution

(a) In class we showed that

$$\overline{X}_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is an unbiased estimator of the mean, for any i.i.d. random variables X_i's.

(b) In class we showed that

$$S_n^2 \triangleq \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1}$$

is an unbiased estimator of the variance, for any i.i.d. random variables X_i's. When X_i's have a normal distribution, we proved (in lecture 2) that $\overline{X}_n$ and S_n^2 are independent.

(c) Recall that when random variables X and Y are independent, the expected value of their product is given by $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$.

Let $T_n \triangleq \overline{X}_n S_n^2$ denote the product of estimators in (a) and (b). Therefore, $\mathbb{E}\{T_n\} = \mathbb{E}\{\overline{X}_n\}\mathbb{E}\{S_n^2\} = \mu^2$. That is, T_n is an unbiased estimator of μ^2.

Note There are several unbiased estimator for μ^2. From Problem Set 1 (Problem 1), we know that the sample mean is a normal random variable, $\overline{X}_n \sim N(\mu, \frac{\mu}{n})$. Thus, the second moment of $(\overline{X}_n)^2$ is

$$\begin{aligned}\mathbb{E}\{(\overline{X}_n)^2\} &\triangleq \mathrm{Var}\{\overline{X}_n\} + \mathbb{E}^2\{\overline{X}_n\} \\ &= \frac{\mu}{n} + \mu^2.\end{aligned}$$

Therefore, by inspection, both estimators,

$$(\overline{X}_n)^2 - \frac{\overline{X}_n}{n} \qquad \text{and} \qquad (\overline{X}_n)^2 - \frac{S_n^2}{n},$$

are unbiased estimators of μ^2.

Problem 3 Let $X_1, X_2, \ldots, X_n$ be i.i.d. Poisson random variables with parameter λ:

$$\mathbb{P}\{X_1 = k\} = \frac{\lambda^k \,\mathrm{e}^{-\lambda}}{k!}, \qquad k = 0, 1, 2, \ldots.$$

(a) Show that $T_n = (\overline{X}_n)^2 - \overline{X}_n$ is a biased estimator of λ^2, find its bias $b_n(\lambda)$, and hence find an unbiased estimator of λ^2. Does $\lim_{n\to\infty} b_n(\lambda) = 0$ for all λ.

(b) Verify that the above estimator is biased using the averaging technique in MATLAB(*ref:* Problem 1, Problem Set 1). Take an appropriate number of realizations of T_n and compute the bias value. Does this agree with your result in part *a*?

Solution

(a) Recall that the first and second moments of a Poisson random variable X_1 with parameter λ are given by

$$\begin{aligned}\mathbb{E}\{X_1\} &= \lambda \\ \mathbb{E}\{X_1^2\} &= \mathrm{Var}\{X_1\} + \mathbb{E}^2\{X_1\} = \lambda + \lambda^2.\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\{T_n\} &= \mathbb{E}\left\{\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)^2\right\} - \mathbb{E}\left\{\overline{X}_n\right\} \\
&= \frac{1}{n^2}\mathbb{E}\left\{\left(\sum_{i=1}^{n} X_i^2\right) + \left(\sum\sum_{1\le i\ne j\le n} X_i X_j\right)\right\} - \mathbb{E}\{X_1\} \\
&= \frac{1}{n^2}\Big(n\mathbb{E}\{X_1^2\} + (n^2-n)\mathbb{E}\{X_1\}\mathbb{E}\{X_2\}\Big) - \lambda \\
&= \lambda^2 - \Big(\frac{n-1}{n}\Big)\lambda.
\end{aligned}$$

Since $\mathbb{E}\{T_n\} \ne \lambda^2$ for some $\lambda > 0$, a statistic T_n is a biased estimator of λ^2. Its bias $b_n(\lambda)$ is given by

$$b_n(\lambda) \triangleq \mathbb{E}\{T_n\} - \lambda^2 = -\Big(\frac{n-1}{n}\Big)\lambda.$$

By inspection, the statistic

$$W_n \triangleq T_n + \Big(\frac{n-1}{n}\Big)\overline{X}_n = (\overline{X}_n)^2 - \frac{\overline{X}_n}{n}$$

is an unbiased estimator of λ since

$$\mathbb{E}\{W_n\} = \mathbb{E}\{T_n\} + \Big(\frac{n-1}{n}\Big)\lambda = \lambda^2,$$

for any $\lambda > 0$.

Note that the limit is non-zero for some λ: $\lim_{n\to\infty} b_n(\lambda) = \lim_{n\to\infty} -(\frac{n-1}{n})\lambda$ $= -\lambda$. That is, T_n is not asymptotically unbiased.

(b) The MATLAB code is given below that generates L realizations of T_n to numerically estimate $\mathbb{E}\{T_n\}$. For the generation of X_i, any value of λ can be selected, code below chooses $\frac{1}{2}$. The ML estimate is a function of sample size n, which can be changed to any desired value. For $n = 1000$ and taking 10000 realizations for estimating expected value of the estimator, the bias is found to be -0.49948610200000. That is quite close to the expected result from part (a).

```
%--------------------------------------------------------------
% Sample size for Xi:
n = 1e3;

% Number of realizations of Tn: that is the averaging length N,
```

```
% as used in Problem 1 of PS 1
L = 1e4;

% Pick an example 'lambda' for the Poisson distribution
lambda = 0.5;

for index = 1:L,
    X = random('poiss',lambda,[n,1]);
    Xbar = sum(X)/n;

    Tn(index) = (Xbar)^2 - Xbar;
end

% check the length of Tn, it should be L;
size(Tn)

% Find E{Tn}, by approximation using averaging. Use this to
% compute the bias value of the estimator [at selected 'n']
bias = sum(Tn)/L - lambda^2

%---------------------------------------------------------------
```

Problem 4 Let X be a binomial $B(n, p)$ random variable with an unknown parameter $0 \le p \le 1$:

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \ldots, n.$$

Find a maximum likelihood estimator of p and compute the estimator variance.

Solution The likelihood function and log-likelihood function are given by

$$L(p) = \binom{n}{X} p^X (1-p)^{n-X}$$
$$\ln L(p) = \ln \binom{n}{X} + X \ln p + (n - X) \ln(1-p),$$

respectively. To derive a MLE for an unknown parameter $0 \le p \le 1$, we consider partial derivatives of the log-likelihood function:

$$\frac{\partial}{\partial p} \ln L(p) = \frac{X}{p} - \frac{n-X}{1-p}$$
$$\frac{\partial^2}{\partial p^2} \ln L(p) = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2}.$$

Setting the first partial derivative to zero and solving for p^*, we have $p^* = X/n$. Notice that the second partial derivative is always negative for any $p \in (0,1)$. Therefore, p^* maximizes the likelihood function. A MLE is then

$$\widehat{p} \triangleq \frac{X}{n}.$$

The variance of this estimator is given by

$$\begin{aligned} \operatorname{Var}\{\widehat{p}\} &= \frac{\operatorname{Var}\{X\}}{n^2} \\ &= \frac{np(1-p)}{n^2} \\ &= \frac{p(1-p)}{n}. \end{aligned}$$

The first equality follows from a property of the variance:

$$\operatorname{Var}\{\alpha Y\} = \alpha^2 \operatorname{Var}\{Y\},$$

for any real number α and any random variable Y. The second equality follows from the expression for the variance of a Binomial $B(n,p)$ random variable.