Probability and Stochastic Processes

A Friendly Introduction for Electrical and Computer Engineers

SECOND EDITION

Roy D. Yates David J. Goodman

Definitions, Theorems, Proofs, Examples, Quizzes, Problems, Solutions

Chapter 7

Section 7.1

Sample Mean: Expected Value and Variance

Definition 7.1 Sample Mean

For iid random variables $X_1, \ldots, X_n$ *with PDF* $f_X(x)$*, the* sample mean *of* X *is the random variable*

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

Theorem 7.1

The sample mean $M_n(X)$ has expected value and variance

$$E\left[M_n(X)\right] = E\left[X\right], \quad Var\left[M_n(X)\right] = \frac{Var\left[X\right]}{n}.$$

Proof: Theorem 7.1

From Definition 7.1, Theorem 6.1 and the fact that $E[X_i] = E[X]$ for all i,

$$E\left[M_n(X)\right] = \frac{1}{n}\left(E\left[X_1\right] + \cdots + E\left[X_n\right]\right) = \frac{1}{n}\left(E\left[X\right] + \cdots + E\left[X\right]\right) = E\left[X\right].$$

Because $\mathrm{Var}[aY] = a^2 \mathrm{Var}[Y]$ for any random variable Y (Theorem 2.14), $\mathrm{Var}[M_n(X)] = \mathrm{Var}[X_1 + \cdots + X_n]/n^2$. Since the X_i are iid, we can use Theorem 6.3 to show

$$\mathrm{Var}[X_1 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] = n\,\mathrm{Var}[X].$$

Thus $\mathrm{Var}[M_n(X)] = n\,\mathrm{Var}[X]/n^2 = \mathrm{Var}[X]/n$.

Section 7.2

Deviation of a Random Variable from the Expected Value

Theorem 7.2 Markov Inequality

For a random variable X such that $P[X < 0] = 0$ and a constant c,

$$P\left[X \geq c^2\right] \leq \frac{E[X]}{c^2}.$$

Proof: Theorem 7.2

Since X is nonnegative, $f_X(x) = 0$ for $x < 0$ and

$$E[X] = \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^{\infty} x f_X(x)\,dx \geq \int_{c^2}^{\infty} x f_X(x)\,dx.$$

Since $x \geq c^2$ in the remaining integral,

$$E[X] \geq c^2 \int_{c^2}^{\infty} f_X(x)\,dx = c^2 P\left[X \geq c^2\right].$$

Revisit

Example 6.18 Problem

If the height X, measured in feet, of a randomly chosen adult is a Gaussian (5.5, 1) random variable, use the Chernoff bound to find an upper bound on $P[X \geq 11]$.

Theorem 6.15 Chernoff Bound

For an arbitrary random variable X and a constant c,

$$P[X \geq c] \leq \min_{s \geq 0} e^{-sc} \phi_X(s).$$

Example 6.18 Solution

In Table 6.1 the MGF of X is

$e^{s\mu + s^2\sigma^2/2}$

$$\phi_X(s) = e^{(11s+s^2)/2}.$$

Thus the Chernoff bound is

$$P[X \geq 11] \leq \min_{s\geq 0} e^{-11s} e^{(11s+s^2)/2} = \min_{s\geq 0} e^{(s^2-11s)/2}.$$

To find the minimizing s, it is sufficient to choose s to minimize $h(s) = s^2 - 11s$. Setting the derivative $dh(s)/ds = 2s - 11 = 0$ yields $s = 5.5$. Applying $s = 5.5$ to the bound yields

$$P[X \geq 11] \leq e^{(s^2-11s)/2}\Big|_{s=5.5} = e^{-(5.5)^2/2} = 2.7 \times 10^{-7}.$$

Based on our model for adult heights, the actual probability (not shown in Table 3.2) is $Q(11 - 5.5) = 1.90 \times 10^{-8}$.

Example 7.1

Let X represent the height (in feet) of a randomly chosen adult. If the expected height is $E[X] = 5.5$, then the Markov inequality states that the probability an adult is at least 11 feet tall satisfies

$$P[X \geq 11] \leq 5.5/11 = 1/2.$$

Theorem 7.3 Chebyshev Inequality

For an arbitrary random variable Y and constant $c > 0$,

$$P[|Y - \mu_Y| \geq c] \leq \frac{\text{Var}[Y]}{c^2}.$$

Proof: Theorem 7.3

In the Markov inequality, Theorem 7.2, let $X = (Y - \mu_Y)^2$. The inequality states

$$P\left[X \geq c^2\right] = P\left[(Y - \mu_Y)^2 \geq c^2\right] \leq \frac{E\left[(Y - \mu_Y)^2\right]}{c^2} = \frac{\text{Var}[Y]}{c^2}.$$

The theorem follows from the fact that $\{(Y - \mu_Y)^2 \geq c^2\} = \{|Y - \mu_Y| \geq c\}$.

Example 7.3 Problem

If the height X of a randomly chosen adult has expected value $E[X] = 5.5$ feet and standard deviation $\sigma_X = 1$ foot, use the Chebyshev inequality to to find an upper bound on $P[X \geq 11]$.

Example 7.3 Solution

Since a height X is nonnegative, the probability that $X \geq 11$ can be written as

$$P[X \geq 11] = P[X - \mu_X \geq 11 - \mu_X] = P[|X - \mu_X| \geq 5.5].$$

Now we use the Chebyshev inequality to obtain

$$P[X \geq 11] = P[|X - \mu_X| \geq 5.5] \leq \text{Var}[X]/(5.5)^2 = 0.033 \approx 1/30.$$

Although this bound is better than the Markov bound, it is also loose. In fact, $P[X \geq 11]$ is orders of magnitude lower than 1/30. Otherwise, we would expect often to see a person over 11 feet tall in a group of 30 or more people!

Section 7.3

Point Estimates of Model Parameters

Definition 7.2 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of the parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

Definition 7.3 Unbiased Estimator

An estimate, $\hat{R}$*, of parameter* r *is* unbiased *if* $E[\hat{R}] = r$*; otherwise,* $\hat{R}$ *is* biased*.*

Asymptotically Unbiased

Definition 7.4 Estimator

The sequence of estimators $\hat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n\to\infty} E[\hat{R}_n] = r.$$

Definition 7.5 Mean Square Error

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = E\left[(\hat{R} - r)^2\right].$$

Theorem 7.4

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \text{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then the sequence $\hat{R}_n$ is consistent.

Proof: Theorem 7.4

Since $E[\hat{R}_n] = r$, we can apply the Chebyshev inequality to $\hat{R}_n$. For any constant $\epsilon > 0$,

$$P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2}.$$

In the limit of large n, we have

$$\lim_{n\to\infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] \leq \lim_{n\to\infty} \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2} = 0.$$

Example 7.4 Problem

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $E[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of r. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

Example 7.4 Solution

First, we observe that $\hat{R}_k$ is an unbiased estimator since

$$E[\hat{R}_k] = E\left[N_k/k\right] = E\left[N_k\right]/k = r.$$

Next, we recall that since N_k is Poisson, $\text{Var}[N_k] = kr$. This implies

$$\text{Var}[\hat{R}_k] = \text{Var}\left[\frac{N_k}{k}\right] = \frac{\text{Var}\left[N_k\right]}{k^2} = \frac{r}{k}.$$

Because $\hat{R}_k$ is unbiased, the mean square error of the estimate is the same as its variance: $e_k = r/k$. In addition, since $\lim_{k\to\infty} \text{Var}[\hat{R}_k] = 0$, the sequence of estimators $\hat{R}_k$ is consistent by Theorem 7.4.

Theorem 7.5

The sample mean $M_n(X)$ is an unbiased estimate of $E[X]$.

Theorem 7.6

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = E\left[(M_n(X) - E[X])^2\right] = \text{Var}[M_n(X)] = \frac{\text{Var}[X]}{n}.$$

Example 7.5 Problem

How many independent trials n are needed to guarantee that $\hat{P}_n(A)$, the relative frequency estimate of $P[A]$, has standard error less than 0.1?

Example 7.5 Solution

Since the indicator X_A has variance $\text{Var}[X_A] = P[A](1 - P[A])$, Theorem 7.6 implies that the mean square error of $M_n(X_A)$ is

$$e_n = \frac{\text{Var}[X]}{n} = \frac{P[A](1 - P[A])}{n}.$$

We need to choose n large enough to guarantee $\sqrt{e_n} \leq 0.1$ or $e_n \leq= 0.01$, even though we don't know $P[A]$. We use the fact that $p(1 - p) \leq 0.25$ for all $0 \leq p \leq 1$. Thus $e_n \leq 0.25/n$. To guarantee $e_n \leq 0.01$, we choose $n = 25$ trials.

Theorem 7.7

If X has finite variance, then the sample mean $M_n(X)$ is a sequence of consistent estimates of $E[X]$.

Proof: Theorem 7.7

By Theorem 7.6, the mean square error of $M_n(X)$ satisfies

$$\lim_{n\to\infty} \operatorname{Var}[M_n(X)] = \lim_{n\to\infty} \frac{\operatorname{Var}[X]}{n} = 0.$$

By Theorem 7.4, the sequence $M_n(X)$ is consistent.

Theorem 7.8 Weak Law of Large Numbers

If X has finite variance, then for any constant $c > 0$,

(a) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| \geq c] = 0$,

(b) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| < c] = 1$.

Theorem 7.9

As $n \to \infty$, the relative frequency $\hat{P}_n(A)$ converges to $P[A]$; for any constant $c > 0$,

$$\lim_{n\to\infty} P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] = 0.$$

Proof: Theorem 7.9

The proof follows from Theorem 7.4 since $\hat{P}_n(A) = M_n(X_A)$ is the sample mean of the indicator X_A, which has mean $E[X_A] = P[A]$ and finite variance $\text{Var}[X_A] = P[A](1 - P[A])$.

Definition 7.6 Convergence in Probability

The random sequence Y_n converges in probability to a constant y if for any $\epsilon > 0$,

$$\lim_{n \to \infty} P\left[|Y_n - y| \geq \epsilon\right] = 0.$$

Definition 7.7 Sample Variance

The sample variance of a set of n independent observations of random variable X is

$$V_n(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_n(X))^2 .$$

Theorem 7.10

$$E\left[V_n(X)\right] = \frac{n-1}{n} Var\left[X\right].$$

Proof: Theorem 7.10

Substituting Definition 7.1 of the sample mean $M_n(X)$ into Definition 7.7 of sample variance and expanding the sums, we derive

$$V_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j.$$

Because the X_i are iid, $E[X_i^2] = E[X^2]$ for all i, and $E[X_i]E[X_j] = \mu_X^2$. By Theorem 4.16(a), $E[X_i X_j] = \text{Cov}[X_i, X_j] + E[X_i]E[X_j]$. Thus, $E[X_i X_j] = \text{Cov}[X_i, X_j] + \mu_X^2$. Combining these facts, the expected value of V_n in Equation (7.22) is

$$\begin{aligned} E[V_n] &= E\left[X^2\right] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\text{Cov}\left[X_i, X_j\right] + \mu_X^2\right) \\ &= \text{Var}[X] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}\left[X_i, X_j\right] \end{aligned}$$

Note that since the double sum has n^2 terms, $\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_X^2 = n^2\mu_X^2$. Of the n^2 covariance terms, there are n terms of the form $\text{Cov}[X_i, X_i] = \text{Var}[X]$, while the remaining covariance terms are all 0 because X_i and X_j are independent for $i \neq j$. This implies

$$E[V_n] = \text{Var}[X] - \frac{1}{n^2}(n\,\text{Var}[X]) = \frac{n-1}{n}\text{Var}[X].$$

Theorem 7.11

The estimate

$$V'_n(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_n(X))^2$$

is an unbiased estimate of $\mathrm{Var}[X]$.

Proof: Theorem 7.11

Using Definition 7.7, we have

$$V_n'(X) = \frac{n}{n-1} V_n(X),$$

and

$$E\left[V_n'(X)\right] = \frac{n}{n-1} E\left[V_n(X)\right] = \text{Var}[X].$$

Quiz 7.3

X is a uniform random variable between -1 and 1 with PDF

$$f_X(x) = \begin{cases} 0.5 & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

What is the mean square error of $V_{100}(X)$, the estimate of $\mathrm{Var}[X]$ based on 100 independent observations of X?

Quiz 7.3 Solution

Define the random variable $W = (X - \mu_X)^2$. Observe that $V_{100}(X) = M_{100}(W)$. By Theorem 7.6, the mean square error is

$$E\left[(M_{100}(W) - \mu_W)^2\right] = \frac{\text{Var}[W]}{100}$$

Observe that $\mu_X = 0$ so that $W = X^2$. Thus,

$$\mu_W = E\left[X^2\right] = \int_{-1}^{1} x^2 f_X(x)\, dx = 1/3$$

$$E\left[W^2\right] = E\left[X^4\right] = \int_{-1}^{1} x^4 f_X(x)\, dx = 1/5$$

Therefore $\text{Var}[W] = E[W^2] - \mu_W^2 = 1/5 - (1/3)^2 = 4/45$ and the mean square error is $4/4500 = 0.000889$.