EE 193 - Applied Probability and Statistics for Engineers
Department of Electrical and Computer Engineering
Tufts University Fall 2007
Problem Set #8
Solutions

Problem 1

Yates and Goodman problems

- **7.1.1**

  (a) The variance of the sample mean is

$$\sigma^2_{M_n X} = \frac{\sigma^2_X}{n}$$

  Since the variance of the exponential is the square of the expected value,

$$\sigma^2_{M_9 X} = \frac{25}{9}$$

  (b) Note that $\lambda = 1/E[X] = 1/5$ for each exponential so that

$$P[X_1 > 7] = \int_7^\infty \frac{1}{5} e^{x/5} dx = e^{-7/5} \approx 0.247$$

  (c) We'll use the central limit theorem.

$$P[M_9 > 7] = P\left[\frac{M_9 - E[M_9]}{\sigma_{M_9}} > \frac{7 - E[M_9]}{\sigma_{M_9}}\right] \approx 1 - \Phi\left(\frac{7 - E[M_9]}{\sigma_{M_9}}\right)$$

  Now $E[M_9] = E[X] = 5$ and $\sigma_{M_9} = \sigma_X/\sqrt{9} = 5/3$. Hence

$$P[M_9 > 7] \approx 1 - \Phi(6/5) \approx 0.1151$$

- **7.1.3** We need the expected value and the variance of Y. First, the expected value:

$$E[Y] = \int_0^1 x^2 dx = 1/3.$$

  Next, the second moment

$$E[Y^2] = \int_0^1 x^4 dx = 1/5.$$

  Thus, the variance is $\sigma^2_Y = 1/5 - (1/3)^2 = 4/45$. The standard error is

$$e_n = \sigma_{M_{50}(X)} = \frac{\sigma_Y}{\sqrt{50}} = \sqrt{\frac{4}{45 \times 50}}$$

- **7.2.4** If we think of rolling the dice as a Bernoulli trial where a success is snake eyes and a failure is anything else then the probability of a success is $p = 1/36$ and $1-p = 35/36$. The RV R is the number of roles until the third success. This is a Pascal RV with $E[R] = 3\times 36 = 108$ and $\sigma_R^2 = 3 \times \frac{35}{36} \times (36)^2 = 3780$.

  (a) According to the Markov inequality

  $$P[R \geq 250] \leq \frac{E[R]}{250} = \frac{108}{250} = 0.432$$

  (b) Since $R > 0$ and $E[R] = 08$

  $$P[R \geq 250] = P[R - E[R] > 250 - E[R]] = P[|R - 108| > 142].$$

  Now, Chebyshev gives

  $$P[|R - 108| > 142] < \frac{\sigma_R^2}{142^2} = 0.1875$$

  (c) Since this is a Pascal PDF, the exact probability is

  $$P[R \geq 250] = 1 - \sum_{r=3}^{249} \binom{r-1}{2} \left(\frac{1}{36}\right)^2 \left(\frac{35}{36}\right)^{r-3}.$$

  Since this sum cannot be done in closed form, in Matlab we use

  ```
  1-pascalpdf(3,1/36,249)
  ```

  to arrive that the answer 0.0299. The bounds are quite loose!

- **7.3.3** Expanding the sum we get

  $$\begin{aligned} E\left[\hat{R}_n\right] &= \frac{1}{n} E\left[X_1(1)X_2(1) + X_1(2)X_2(2) + \cdots + X_1 n X_2(n)\right] \\ &= \frac{1}{n}\left(E[X_1(1)X_2(1)] + E[X_1(2)X_2(2)] + \ldots E[X_1(n)X_2(n)]\right)/. \end{aligned}$$

  Since the trials are independent and identically distributed, all of the terms in the parentheses are the same and equal to $E[X_1X_2] = r$, the correlation. Hence

  $$E\left[\hat{R}_n\right] = \frac{1}{n} \times nr = r$$

  and we conclude that the estimator is not biased.

  Now, we need the find the variance of $\hat{R}_n$. To do this easily, we define the random variable $Y_i = X_1(i)X_2(i)$ so that we can write $R_n = 1/n \sum_{i=1}^n Y_i$. Note that since the X_i's are independent and identically distributed from one i to the next, so too are the Y_i. Hence, from basic theory about sums of IID random variable, $\text{Var}[\hat{R}_n] = \text{Var}[Y]/n$. Noting that $\text{Var}[Y] = \text{Var}[X_1X_2]$ we see that as n goes to infinity, the variance of $\hat{R}_n$ goes to zero so we conclude that the estimator is consistent.

- **7.4.1**

(a) We have $P_X(1) = 0.9$ and since X is Bernoulli, $E[X] = 0.9$. Hence, the two are the same.

(b) This is a straightforward application of the equation in part (a) of Theorem 7.12 on page 286 of the text. Plugging in the specifics of this problem to that equation yields:

$$P[|M_{90}(X) - \mu_X| \geq 0.05] \leq \frac{\text{Var}[X]}{90 \times (0.05)^2} = \alpha$$

Since X is a Bernoulli, $\text{Var}[X] = 0.9 \times 0.1$ and we have after a bit of calculation that $\alpha = 0.4$. Thus we have concluded that the probability that with 90 flips we are off from p by more that 0.05 is less than 40%.

(c) Again, this is an application of the same equation which in this case asks that we find n to satisfy

$$0.1 = \frac{0.9 \times 0.1}{n(0.03)^2} \rightarrow n = 100.$$

- **7.4.6** This problem follows the same ideas as in Example 7.6 on page 287 of the text. In that example, we have

$$P\left[|\hat{P}_n(A) - P[A]| < c\right] \geq 1 - \frac{P[A](1 - P[A])}{nc^2}$$

We apply this equation in two different ways to answer the question.

(a) Here we are given $c = 0.001$, $P[A] = 0.9$ and $1 - \alpha = 0.99$ and are being asked for an n such that the right hand side of the above equation equals $1 - \alpha$

$$1 - \frac{0.01 \times 0.99}{n(0.001)^2} = 0.99 \rightarrow n \geq 990,000$$

(b) In this case the mathematical statement of the problem is to find an n such that

$$P\left[|\hat{P}_n(A) - P[A]| < 0.1 \times 0.01 \times P[A]\right] \geq 1 - \frac{P[A](1 - P[A])}{n(0.1 \times 0.01 \times P[A])^2} = 0.99$$

Again, since we have $P[A] = 0.01$ we solve for n to obtain $n \geq 9.9 \times 10^9$.

- **8.1.2**

(a) We are looking to find c that satisfies

$$P[|K - E[K]| > c] = 0.05 \quad \text{or} \quad P[|K - E[K]| < c] = 0.95.$$

Now, the null hypothesis is that the coin is fair meaning $p = 0.5$. Under this hypothesis, K will be binomial with $\mu_k = 100 \times 0.5 = 50$ and $\sigma_K^2 = 100 \times 0.5 \times (1 - 0.5) = 25$. To find c, we make use of the central limit theorem:

$$P\left[\frac{|K - E[K]}{\sigma_K} < \frac{c}{\sigma_K}\right] \approx 2\Phi\left(\frac{c}{\sigma_K}\right) - 1 = 0.95.$$

Since $\sigma_K = 5$ we can solve for c and get $c > 9.8$. In other words, we can reject the hypothesis that $p = 0.5$ with confidence level 0.95 is we observe more than $50 + 10 = 60$ or fewer than $50 - 10 = 40$ heads in 100 flips.

(a) Now we are looking for a c' that satisfies

$$P[K > c'] = 0.01.$$

We start by standardizing

$$P\left[\frac{K - E[K]}{\sigma_K} > \frac{c' - E[K]}{\sigma_K}\right] = 0.01$$

and then using the central limit theorem

$$P\left[\frac{K - E[K]}{\sigma_K} > \frac{c' - E[K]}{\sigma_K}\right] \approx 1 - \Phi\left(\frac{c' - 50}{5}\right) = 0.01$$

Solving for c' yields $c' \approx 61.75$. So if we see 62 or more heads, we should reject the null hypothesis of a fair coin.

- **8.1.3**

(a) At a temperature of 100°C, the null hypothesis is that the lifetime of a chip is an exponential with $E[X] = (200/100)^2 = 4$ years. Let A be the event that a chip fails in one day. Assuming that every year has 365 days (i.e., ignoring leap years), the probability of a given chip failing in one day is

$$P[A] = \int_0^{1/365} 1/4e^{-x/4}dx = 1 - e^{-1/1460}$$

Now, in words, the problem asks us to find the probability that we see more than $N = 0$ failures from m tested chips given that the hypothesis is true. This probability is the significance level. Since each chip is (assumed to be) independent of the rest, we have a Bernoulli process and the number of failed chips in m trials is a binomial with probability of "success" being $P[A]$ calculated in the above equation. Hence

$$P_N(n) = \binom{m}{n}(P[A])^n(1 - P[A])^{m-n}$$

and $P[N > 0] = 1 - P[N = 0] = 1 - (1 - P[A])^m = 1 - e^{m/1460} = \alpha$

(b) Here we are given $\alpha = 0.01$ so we can solve for m to obtain $m > 14.74$. So we need to test 15 chips.

(c) Here need to compute $P[A]$ as a function of temperature. As in part (a), this is

$$P[A] = \int_0^{1/365} \left(\frac{T}{200}\right)^2 e^{-\left(\frac{T}{200}\right)^2 x}dx = 1 - e^{-\frac{T^2}{1.46\times10^7}}.$$

The equation for α is now

$$1 - e^{-\frac{mT^2}{1.46\times10^7}} = \alpha \rightarrow m = -\frac{1}{T^2} \times 1.46 \times 10^7 \times \ln(1 - \alpha).$$

So, as T increases, m decreases. This is due to the fact that as T increases, the average failure times decreases, so with 1 day of testing, it is more likely we will see failed chips. Hence, we need to test fewer of them.