Homework based on Chapter 17, 19
Computational Probability and Statistics
CIS 2033, Section 002

Due: 9:00 AM, Friday, April 17, 2015

17.4 During the Second World War, London was hit by numerous flying bombs. The following data are from an area in South London of 36 square kilometers. The area was divided into 576 squares with sides of length 1/4 kilometer. For each of the 576 squares the number of hits was recorded. In this way we obtain a dataset $x_1, x_2, \ldots, x_{576}$, where x_i denotes the number of hits in the ith square. The data are summarized in the following table which lists the number of squares with no hits, 1 hit, 2 hits, etc.

Number of hits	0	1	2	3	4	5	6	7
Number of squares	229	211	93	35	7	0	0	1

Source: R.D. Clarke. An application of the Poisson distribution. *Journal of the Institute of Actuaries*, 72:48, 1946; Table 1 on page 481. © Faculty and Institute of Actuaries.

An interesting question is whether London was hit in a completely random manner. In that case a Poisson distribution should fit the data.

a. If we model the dataset as the realization of a random sample from a Poisson distribution with parameter μ, then what would you choose as an estimate for μ?

b. Check the fit with a Poisson distribution by comparing some of the observed relative frequencies of 0's, 1's, 2's, etc., with the corresponding probabilities for the Poisson distribution with μ estimated as in part **a**.

17.4

a. Let the random variable X denote the number of hits and $X \sim Pois(\mu)$. Then $E[X] = \mu$.

$$\begin{aligned}\mu &= \frac{0 \times 229 + 1 \times 211 + 2 \times 93 + 3 \times 35 + 4 \times 7 + 5 \times 0 + 6 \times 0 + 7 \times 1}{576} \\ &= \frac{537}{576} = 0.9323\end{aligned}$$

b. We compute the frequency by using freq$(k) = \frac{\#squaresinkhits}{\#totalsquares{:}576}$ and compute the probability by using $P(X = k) = \frac{\mu^k}{k!}e^{-\mu}$, for $k = 0, 1, \ldots, 7$ where $\mu = 0.9323$,

# hits	0	1	2	3	4	5	6	7
# squares	229	211	93	35	7	0	0	1
freq $= \frac{\#squares}{\#totalsquares:576}$	0.3976	0.3663	0.1615	0.0608	0.0122	0	0	0.0017
$P(X=k) = \frac{\mu^k}{k!}e^{-\mu}$	0.3936	0.3670	0.1711	0.0532	0.0124	0.0023	0.0004	0.0000

17.6 Recall Exercise 15.1 about the chest circumference of 5732 Scottish soldiers, where we constructed the histogram displayed in Figure 17.11. The histogram suggests modeling the data as the realization of a random sample from a normal distribution.

a. Suppose that for the dataset $\sum x_i = 228377.2$ and $\sum x_i^2 = 9124064$. What would you choose as estimates for the parameters μ and σ of the $N(\mu, \sigma^2)$ distribution?
Hint: you may want to use the relation from Exercise 16.15.

b. Give an estimate for the probability that a Scottish soldier has a chest circumference between 38.5 and 42.5 inches.

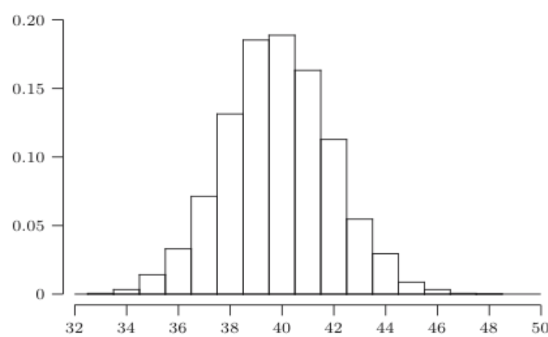

Fig. 17.11. Histogram of chest circumferences.

17.6

a. $N(\mu, \sigma^2)$, we can

$$\mu = \frac{1}{n}\sum_i x_i = \frac{1}{5732}228377.2 = 39.8425$$

$$\begin{aligned}\sigma &= \sqrt{\frac{1}{n-1}\sum_i (x_i - \mu)^2} \\ &= \sqrt{\frac{1}{n-1}\sum_i (x_i^2 + \mu^2 - 2\mu x_i)} \\ &= \sqrt{\frac{1}{n-1}\sum_i x_i^2 + \sum_i \mu^2 - \sum_i 2\mu x_i} \\ &= \sqrt{\frac{1}{n-1}\sum_i x_i^2 + n\mu^2 - 2\mu \sum_i x_i} \\ &= \sqrt{\frac{1}{5732-1}9124064 + 5732(39.8425)^2 - 2(39.8425)228377.2} \\ &= 2.0863\end{aligned}$$

b. Given $X \sim N(\mu, sigma^2)$, where $\mu = 39.8425, \sigma^2 = 4.3526$, let $Z = \frac{X-\mu}{\sigma}$, and Z is a standard norm distribution such that $Z \sim N(0,1)$. We have to compute

$$\begin{aligned}P(38.5 < X < 42.5) &= P(X < 42.5) - P(X < 38.5) \\ &= P(Z < \frac{42.5-\mu}{\sigma}) - P(Z < \frac{38.5-\mu}{\sigma}) \\ &= \Phi(\frac{42.5-\mu}{\sigma}) - \Phi(\frac{38.5-\mu}{\sigma}) \\ &= \Phi(1.2738) - \Phi(-0.6435) \\ &= 0.8980 - 0.2611 \\ &= 0.6369\end{aligned}$$

Note that $\Phi(1.2738) = P(Z \leq 1.2738) = 1 - \hat{p}$, where $\hat{p}$ is the value obtained from the lookup table; $\Phi(-0.6435) = P(Z \leq -0.6435) = P(Z \geq 0.6435) = \tilde{p}$ where $\tilde{p}$ is the valued obtained from the lookup table.

19.2 Suppose the random variables $X_1, X_2, \ldots, X_n$ have the same expectation μ.

a. Is $S = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$ an unbiased estimator for μ ?

b. Under what conditions on constants $a_1, a_2, \ldots, a_n$ is $T = a_1X_1 + a_2X_2 + \ldots a_nX_n$ an unbiased estimator for μ ?

For a), $E[S] = E[\frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3] = \frac{1}{2}E[X_1] + \frac{1}{3}E[X_2] + \frac{1}{6}E[X_3] = \frac{1}{2}\mu + \frac{1}{3}\mu + \frac{1}{6}\mu = \mu$. So, S is unbiased estimator for μ.

For b), T is an unbiased estimator for μ, then $E[T] = \mu$. Then, $E[T] = E[a_1X_1 + a_2X_2 + \ldots a_nX_n] = a_1E[X_1] + a_2E[X_2] + \ldots + a_nE[X_n] = a_1\mu + a_2\mu + ldots + a_n\mu + (a_1 + a_2 + \ldots + a_n)\,\mu = \mu$, then $a_1 + a_2 + \ldots + a_n = 1$.

19.5 Suppose a dataset is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from an $Exp(\lambda)$ distribution, where $\lambda > 0$ is unknown. Let μ denote the corresponding expectation and let M_n denote the minimum of $X_1, X_2, \ldots, X_n$. Recall that M_n has an $Exp(n\lambda)$ distribution. Find out for which constant c the estimator $T = cM_n$ is an unbiased estimator for μ.

We know that $E[M_n] = \frac{1}{n\lambda}$ since $M_n \sim Exp(n\lambda)$.
Then $E[T] = E[cM_n] = cE[M_n] = \frac{c}{n\lambda}$. Let T be an unbiased estimator for μ, then $E[T] = \mu$, then $\frac{c}{n\lambda} = \mu$, then $c = n\lambda\mu$. Moreover, from the question, we know that μ is the expectation of a $Exp(\lambda)$, then $\mu = \frac{1}{\lambda}$. Then, $c = n$.

19.7 Leaves are divided into four different types: starchy-green, sugary-white, starchy-white, and sugary-green. According to genetic theory, the types occur with probabilities $\frac{1}{4}(\theta+2)$, $\frac{1}{4}\theta$, $\frac{1}{4}(1-\theta)$, and $\frac{1}{4}(1-\theta)$, respectively, where $0<\theta<1$. Suppose one has n leaves. Then the number of starchy-green leaves is modeled by a random variable N_1 with a $Bin(n,p_1)$ distribution, where $p_1=\frac{1}{4}(\theta+2)$, and the number of sugary-white leaves is modeled by a random variable N_2 with a $Bin(n,p_2)$ distribution, where $p_2=\frac{1}{4}\theta$. The following table lists the counts for the progeny of self-fertilized heterozygotes among 3839 leaves.

Type	Count
Starchy-green	1997
Sugary-white	32
Starchy-white	906
Sugary-green	904

Source: R.A. Fisher. *Statistical methods for research workers.* Hafner, New York, 1958; Table 62 on page 299.

Consider the following two estimators for θ:

$$T_1 = \frac{4}{n}N_1 - 2 \quad \text{and} \quad T_2 = \frac{4}{n}N_2.$$

19.7

a. Check that both T_1 and T_2 are unbiased estimator for θ.

b. Compute the value of both estimators for θ.

For a), $E[T_1]=E[\frac{4}{n}N_1-2]=\frac{4}{n}E[N_1]-2=\frac{4}{n}np_1-2=\frac{4}{n}n\frac{1}{4}(\theta+2)-2=\theta+2-2=\theta$. So, T_1 is an unbiased estimator for θ.

$E[T_2]=E[\frac{4}{n}N_2]=\frac{4}{n}E[N_2]=\frac{4}{n}np_2=\frac{4}{n}n\frac{1}{4}\theta=\theta$. So, T_2 is an unbiased estimator for θ.

b), $n=3839, n_1=1997, n_2=32, E[N_1]=1997, E[N_2]=32$, $E[T_1]=\frac{4}{n}E[N_1]-2=\frac{4}{3839}1997-2=0.0808$.
$E[T_2]=\frac{4}{n}E[N_2]=\frac{4}{3839}32=0.0333$.