Stat 345/445: Theoretical Statistics I: Homework 8 Solutions

Textbook Exercises

5.3 (345: 1 pt.) Let $X_1, \dots, X_n$ be iid random variables with continuous cdf F_X, and suppose $\mathbb{E}X_i = \mu$. Define the random variables $Y_1, \dots, Y_n$ by

$$Y_i = \begin{cases} 1 & \text{if } X_i > \mu \\ 0 & \text{if } X_i \leq \mu. \end{cases}$$

Find the distribution of $\sum_{i=1}^n Y_i$.

Note that $Y_i \sim$ Bernoulli with $p_i = P(X_i \geq \mu) = 1 - F(\mu)$ for each i.
Since the Y_i's are iid Bernoulli,

$$\sum_{i=1}^n Y_i \sim \text{binomial}(n, p = 1 - F(\mu)).$$

5.4 (445: 2 pts.) A generalization of iif random variables is *exchangeable* random variables, an idea due to deFinetti (1972). A discussion of exchangeability can also be found in Feller (1971). The random variables $X_1, \dots, X_n$ are *exchangeable* if any permutation of any subset of them of size k ($k \leq n$) has the same distribution. In this exercise we will see an example of random variables that are exchangeable but not iid. Let $X_i|P \sum$ iid Bernouli(P), $i = 1, \dots, n$, and let $P \sim$ uniform (0,1).

(a) Show that the marginal distribution of any k of the Xs is the same as

$$P(X_1 = x_1, \dots, X_k = x_k) = \int_0^1 p^t(1-p)^{k-t}dp = \frac{t!(k-t)!}{(k+1)!},$$

where $t = \sum_{i=1}^k x_i$. Hence, the Xs are exchangeable.

$$X_i|P \sim \text{Bernoulli}(P). \qquad P \sim \text{uniform}(0,1).$$

$$P(X_1 = x_1, \dots, X_k = x_k|P) = P^{\sum_{i=1}^k x_i}(1-p)^{k-\sum_{i=1}^k x_i} = p^t(1-p)^{k-t}$$

where $t = \sum_{i=1}^k x_i$.

$$\begin{aligned} P(X_1 = x_1, \dots, X_k = x_k) &= \int_{-\infty}^{\infty} P(X_1 = x_1, \dots, X_k = x_k|P)f(p)dp \\ &= \int_0^1 p^t(1-p)^{k-t}dp \\ &= \beta(t+1, k-t+1) \\ &= \frac{\Gamma(t+1)\Gamma(k-t+1)}{\Gamma(k+2)} \\ &= \frac{t!(k-t)!}{(k+1)!} \end{aligned}$$

(b) Show that, marginally,

$$P(X_1 = x_1, \dots, X_n = x_n) \neq \Pi_{i=1}^n P(X_i = x_i),$$

so the distribution of the Xs is exchangeable but not iid.

$$P(X_i = x_i) = \frac{x_i!(1-xi)!}{2!} = \begin{cases} \frac{1}{2} & x_i = 0 \\ \frac{1}{2} & x_i = 1 \end{cases}$$

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = (\frac{1}{2})^n \neq \frac{t!(n-t)!}{(n+1)!} \implies \text{not iid}$$

5.6 If X has pdf $f_X(x)$ and Y, independent of X, has pdf $f_Y(y)$, establish formulas, similar to (5.2.3), for the random variables Z in each of the following situations.

(a) (345 & 445: 1 pt.) $Z = X - Y$

For $Z = X - Y$, set $W = X$. Then $Y = W - Z$, $X = W$, and $|J| = \begin{vmatrix} 0 & 1 \\ -1 & 1 \end{vmatrix} = 1$. Then $f_{Z,W}(z, w) = f_X(w) f_Y(w - z) \cdot 1$, thus $f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(w - z) dw$.

(b) (345 & 445: 1 pt.) $Z = XY$

For $Z = XY$, set $W = X$. Then $Y = Z/W$ and $|J| = \begin{vmatrix} 0 & 1 \\ 1/w & -z/w^2 \end{vmatrix} = -1/w$. Then $f_{Z,W}(z, w) = f_X(w) f_Y(z/w) \cdot |-1/w|$, thus $f_Z(z) = \int_{-\infty}^{\infty} |-1/w| f_X(w) f_Y(z/w) dw$.

5.10 Let $X_1, \dots, X_n$ be a random sample from a $n(\mu, \sigma^2)$ population.

(a) (345 & 445: 1.5 pts.) Find expressions for $\theta_1, \dots, \theta_4$, as defined in Exercise 5.8, in terms of μ and σ^2.

$$\begin{aligned}
\theta_1 &= \mathbb{E}X_i && = \mu \\
\theta_2 &= \mathbb{E}(X_i - \mu)^2 && = \sigma^2 \\
\theta_3 &= \mathbb{E}(X_i - \mu)^3 && \\
&= \mathbb{E}(X_i - \mu)^2 (X_i - \mu) && \text{(Stein's lemma: } \mathbb{E}g(X)(X - \theta) = \sigma^2 \mathbb{E}g'(X)) \\
&= 2\sigma^2 \mathbb{E}(X_i - \mu) && = 0 \\
\theta_4 &= \mathbb{E}(X_i - \mu)^4 && \\
&= \mathbb{E}(X_i - \mu)^3 (X_i - \mu) && \\
&= 3\sigma^2 \mathbb{E}(X_i - \mu)^2 && = 3\sigma^4
\end{aligned}$$

(b) (345: 1 pt. & 445: 0.5 pt.) Use the results from Exercise 5.8, together with the results of part (a), to calculate Var S^2.

$$\text{Var}S^2 = \frac{1}{n}(\theta_4 - \frac{n-3}{n-1}\theta_2^2) = \frac{1}{n}(3\sigma^4 - \frac{n-3}{n-1}\sigma^4) = \frac{2\sigma^4}{n-1}.$$

(c) (345: 1 pt. & 445: 0.5 pt.) Calculate Var S^2 a completely different (and easier) way: Use the fact that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

Use the fact that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ and Var$\chi^2_{n-1} = 2(n-1)$ to get

$$\text{Var}\Big(\frac{(n-1)S^2}{\sigma^2}\Big) = 2(n-1)$$

which implies $(\frac{(n-1)^2}{\sigma^4})\text{Var}S^2 = 2(n-1)$ and hence

$$\text{Var}S^2 = \frac{2(n-1)}{(n-1)^2/\sigma^4} = \frac{2\sigma^4}{n-1}.$$

5.15 (345 & 445: 2 pts.) Establish the following recursion relations for means and variances. Let $\bar{X}_n$ and S_n^2 be the mean and variance, respectively, of $X_1, \ldots, X_n$. Then suppose another observation, X_{n+1}, becomes available. Show that

(a) $\bar{X}_{n+1} = \frac{X_{n+1}+n\bar{X}_n}{n+1}$.

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1} = \frac{X_{n+1} + \sum_{i=1}^{n} X_i}{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}.$$

(b) $nS_{n+1}^2 = (n-1)S_n^2 + (\frac{n}{n+1})(X_{n+1} - \bar{X}_n)^2$.

$$\begin{aligned}
nS_{n+1}^2 &= \frac{n}{(n+1)-1}\sum_{i=1}^{n+1}(X_i - \bar{X}_{n+1})^2 \\
&= \sum_{i=1}^{n+1}(X_i - \frac{X_{n+1}+n\bar{X}_n}{n+1})^2 && \text{use (a)} \\
&= \sum_{i=1}^{n+1}(X_i - \frac{X_{n+1}}{n+1} - \frac{n\bar{X}_n}{n+1})^2 \\
&= \sum_{i=1}^{n+1}\Big[(X_i - \bar{X}_n) - \Big(\frac{X_{n+1}}{n+1} - \frac{\bar{X}_n}{n+1}\Big)\Big]^2 \\
&= \sum_{i=1}^{n+1}\Big[(X_i - \bar{X}_n)^2 - 2(X_i - \bar{X}_n)\Big(\frac{X_{n+1} - \bar{X}_n}{n+1}\Big) + \frac{1}{(n+1)^2}\Big(X_{n+1} - \bar{X}_n\Big)^2\Big] \\
&= \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + (X_{n+1} - \bar{X}_n)^2 - 2\frac{(X_{n+1} - \bar{X}_n)^2}{n+1} + \frac{n+1}{(n+1)^2}(X_{n+1} - \bar{X}_n)^2 \\
&= (n-1)S_n^2 + \frac{n}{n+1}(X_{n+1} - \bar{X}_n)^2.
\end{aligned}$$

Since $\sum_1^n (X_i - \bar{X}_n) = 0$.

5.16 (345 & 445: 1.5 pts.) Let X_i, $i = 1, 2, 3$, be independent with $n(i, i^2)$ distributions. For each of the following situations, use the X_is to construct a statistic with the indicated distribution.

(a) chi squared with 3 degrees of freedom

$$\sum_{i=1}^{3}(\frac{X_i - i}{i})^2 \sim \chi_3^2$$

(b) t distributions with 2 degrees of freedom

$$(\frac{X_i - 1}{i})\Big/\sqrt{\sum_{i=2}^{3}(\frac{X_i - i}{i})^2\Big/2} \sim t_2$$

(c) F distribution with 1 and 2 degrees of freedom

Square the random variable in part (b).