# SOLUTION FOR HOMEWORK 1, STAT 6331

Well, welcome to your first homework. In my solutions you may find some seeded mistakes (this is why it is not a good idea just to copy my solution. If you find them — please, do not e-mail or call me. Instead, write down them on the first page of your solutions and you may give you some partial credit for them — but keep in mind that the total for your homeworks cannot exceed 20 points.

Now let us look at your problems.

1. Problem 5.6. (page 256). Given: $X \sim f_X(x)$, $Y \sim f_Y(y)$, the two RVs are independent and, while it is not specified, they are supported on a real line.

a). Find the pdf of $Z = X - Y$. The answer is $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x-z)dx = \int_{-\infty}^{\infty} f_Y(y) f_X(z+y)dy$, which is "obvious" to me, but let us consider several possible methods of solution.

Method A. I use the bivariate transformation method (see Section 4.3). Consider a new system of two (one-to-one) random variables

$$\begin{cases} Z = X - Y \\ W = X. \end{cases}$$

Solve it with respect to the original random variables and get

$$\begin{cases} X = W \\ Y = W - Z. \end{cases}$$

For the last system the Jacobian is

$$J = \left| \begin{matrix} 0 & 1 \\ -1 & 1 \end{matrix} \right] = 1.$$

This yields the following joint density

$$f_{Z,W}(z, w) = |J| f_X(w) f_Y(w - z).$$

We need the marginal pdf of $Z$, so let us integrate the bivariate,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z, w)dw = \int_{-\infty}^{\infty} f_X(w) f_Y(w - z)dw.$$

Method B. Here we use the more universal method of differentiating a corresponding cdf. Write,

$$F_Z(z) = \mathrm{P}(X - Y \le z) = \int_{(x,y):\ x-y\le z} f_X(x) f_Y(y)dydx$$

$$= \int_{-\infty}^{\infty} f_X(x)[\int_{y\ge x-z} f_Y(y)dy]dx = \int_{-\infty}^{\infty} f_X(x)[1 - F_Y(x - z)]dx.$$

Now we need to differentiate the cdf,

$$f_Z(z) = dF_Z(z)/dz = \int_{-\infty}^{\infty} f_X(x) f_Y(x-z)dx.$$

Method C. OK, we know that $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-y)dy$. Let us use it and get the wished pdf. Introduce $U = -Y$ and note that $f_U(u) = f_Y(-u)$ (you should know how to prove the last relation). Then $Z := X - Y = X + U$ and

$$f_Z(z) = f_{X+U}(z) = \int_{-\infty}^{\infty} f_X(x) f_U(z-x)dx = \int_{-\infty}^{\infty} f_X(x) f_Y(x-z)dx.$$

b) Find the pdf of $Z := XY$.

Solution: Let us see how Method A will work out. Introduce a new pair of one-to-one RVs

$$\begin{cases} Z = XY \\ W = X. \end{cases}$$

Solve it with respect to the old variables

$$\begin{cases} Y = Z/W \\ X = W. \end{cases}$$

The corresponding Jacobian is

$$J = \left| \begin{array}{cc} 0 & 1 \\ 1/w & -z/w^2 \end{array} \right] = -\frac{1}{w}.$$

Using this we get

$$f_{Z,W}(z,w) = |-w^{-1}| f_X(w) f_Y(z/w),$$

and taking the integral over $w$ we get the desired pdf,

$$f_Z(z) = \int_{-\infty}^{\infty} |-x^{-1}| f_X(x) f_Y(z/x)dx.$$

Just for fun, let us check Method B here. Write,

$$F_Z(z) = \mathrm{P}(XY \le z) = \int_{(x,y):\, xy \le z} f_X(x) f_Y(y)dy$$

$$= \int_{-\infty}^{\infty} f_X(x)[\int_{y\ \{(y \le z/x)\cap(x\ge 0)\}\cup\{(y>z/x)\cap(x<0)\}} f_Y(y)dy]dx$$

$$= \int_0^{\infty} f_X(x) F_Y(z/x)dx + \int_{-\infty}^{0} f_X(x)[1 - F_Y(z/x)]dx.$$

Now we take the derivative and get the answer,

$$f_Z(z) = dF_Z(z)/dz = \int_0^{\infty} f_X(x) x^{-1} f_Y(z/x)dx + \int_{-\infty}^{0} f_X(x)(-1)x^{-1} f_Y(z/x)dx$$

$$= \int_{-\infty}^{\infty} |x|^{-1} f_X(x) f_Y(z/x) dx.$$

c) Find the pdf of $Z = X/Y$. Let us check Method A. Write Introduce a new pair of one-to-one RVs

$$\begin{cases} Z = X/Y \\ W = X. \end{cases}$$

Solve it with respect to the old variables

$$\begin{cases} Y = W/Z \\ X = W. \end{cases}$$

The corresponding Jacobian is

$$J = \left| \begin{array}{cc} 1/z & -w/z^2 \\ 1 & 0 \end{array} \right] = \frac{w}{z^2}.$$

Using this we get

$$f_{Z,W}(z,w) = |w/z^2| f_X(w) f_Y(w/z),$$

and taking the integral over $w$ we get the desired pdf,

$$f_Z(z) = z^{-2} \int_{-\infty}^{\infty} |x| f_X(x) f_Y(x/z) dx.$$

2. Problem 5.11. Jensen's inequality tells us that if $g(x)$ is convex (it has a positive second derivative with $x^2$ being an example) then $E\{g(Y)\} \geq g(E\{Y\})$ with the equality iff for every line $a + bx$, which is tangent to $g(x)$ at $x = E\{X\}$, we have $P(g(X) = a + bX) = 1$.

Using Jensen's inequality we get that for the function $g(s) = s^2$, which is convex, we can write

$$\sigma^2 = E\{S^2\} \geq [E\{S\}]^2.$$

This yields $ES \leq \sigma$. If $\sigma = 0$ then the random variable is degenerated and takes on a single value, and we have equality. Otherwise, because $s^2$ is not a linear function, we have $ES < \sigma$.

Remark: for this simple function we can use the following familiar relation $|EX|^2 \leq EX^2$ which follows from $E\{(X - E\{X\})^2 \geq 0$, and you can also see when the equality takes place. Another useful application of Jensen's inequality is that for positive random variables

$$E\{1/X\} \geq 1/E\{X.$$

3. Problem 5.16. Well, this is a good problem where we remember the classical RV's and how to create them. In what follows, $X_i$, $i = 1, 2, 3$ are $N(i, i^2)$ (meaning normally distributed with mean $i$ and variance $i^2$. Further, denote by $Z_k$, $k = 0, 1, 2, \ldots$ iid standard normal RVs.

a) Construct $\chi_3^2$. Remember that chi-squared RV with 3 degrees of freedom can be defined as

$$\chi_3^2 \stackrel{D}{=} Z_1^2 + Z_2^2 + Z_3^2,$$

where $\stackrel{D}{=}$ means equality in distribution, that is, the left random variable has the same distribution as the random variable on the right. In words, a chi-squared random variable with $k$ degrees of freedom has the same distribution as the sum of $k$ squared iid standard normal RVs. Further, $(X_i - i)/i \stackrel{D}{=} Z_0$ according to the property of a normal random variable (remember z-scoring). As a result,

$$\chi_3^2 \stackrel{D}{=} \sum_{i=1}^{3}[(X_i - i)/i]^2.$$

b) A random variable $T_k$ has t-distribution with $k$ degrees of freedom iff

$$T_k \stackrel{D}{=} \frac{Z_0}{[\chi_k^2/k]^{1/2}}.$$

As a result,

$$T_2 \stackrel{D}{=} \frac{X_1 - 1}{[\sum_{i=2}^{3}[(X_i - i)/i]^2/2]^{1/2}}.$$

c) By definition

$$F_{p,q} \stackrel{D}{=} \frac{\chi_p^2/p}{\tilde{\chi}_q^2/q},$$

where $\chi_p^2$ and $\chi_q^2$ are independent. Using this we get

$$F_{1,2} \stackrel{D}{=} \frac{(X_1 - 1)^2}{\sum_{i=2}^{3}[(X_i - i)/i]^2/2}.$$

4. Problem 5.21. Here you should remember that whenever the maximum of several random variables is studied and the question is that the maximum is larger than something, it is easier to work with the complementary event that the maximum is smaller or equal to...

Let us begin the solution. Let $X$ and $Y$ be iid from a distribution with median $m$. Then the probability in question is

$$\mathrm{P}(\max(X,Y) > m) = 1 - \mathrm{P}(\max(X,Y) \le m)$$

$$= 1 - \mathrm{P}(X \le m)\mathrm{P}(Y \le m) = 1 - (1/2)(1/2) = 3/4.$$

What a surprising result! Now, if I may, let us look at a general case for $X_{(n)} := \max(X_1, X_2, \ldots, X_n)$.

$$\mathrm{P}(X_{(n)} > m) = 1 - \mathrm{P}(X_{(n)} \le m) = 1 - \prod_{l=1}^{n} \mathrm{P}(X_l \le m) = 1 - 2^{-n}.$$

5. Problem 5.29. This is the problem where the CLT should be used. It states that for iid RVs $(X_1, X_2, \ldots, X_n)$ with finite mean $\mu$ and variance $\sigma^2$ (our case) we have $n^{1/2}(\bar{X} - \mu)/\sigma \stackrel{D}{\to} N(0,1)$ as $n \to \infty$. Remember that the rule of thumb is that for $n \ge 30$ the normal approximation can be used for all practical purposes.

Now the solution. We have $\mu = 1$, $\sigma = 0.05$, $n = 100$, and let $Z$ be a standard normal RV. Write

$$\mathrm{P}(\text{ 100 booklets weigh more that 100.4 }) = \mathrm{P}(\sum_{l=1}^{n} X_l > 100.4) = \mathrm{P}(\bar{X} > 1.004)$$

$$= \mathrm{P}(n^{1/2}\frac{\bar{X}-\mu}{\sigma} > n^{1/2}\frac{1.004-m}{\sigma}) = P(Z > n^{1/2}\frac{1.004-\mu}{\sigma})$$
$$= P(Z > 10(1.004-1)/0.05) = \mathrm{P}(Z > 0.8) = 0.2119.$$

In the last equality I used Table.

Remark: Do you see $\sigma_{\bar{X}} = \sigma/n^{1/2}$ everywhere? Remember that the classical z-scoring is $(\bar{X}-\mu)/\sigma_{\bar{X}}$ which yields a standard normal RV (for large $n$). We shall discuss this in detail shortly in Problem 5.34.

6. Problem 5.31. This problem, as you probably already realized, is about Margin of Error (M). Specifically, here $n = 100$ and $\sigma^2 = 9$. Given

$$\mathrm{P}(|\bar{X}-\mu| \le M) = .9,$$

we are asked to find the margin of error $M$. Note that $1-\alpha = 0.9$ is the confidence coefficient.

a) Solution via Chebyshev's inequality. Write

$$0.9 = \mathrm{P}(|\bar{X}-\mu| \le M) = 1 - \mathrm{P}(|\bar{X}-\mu| > E)$$

$$\ge 1 - \mathrm{Var}(\bar{X})/E^2 = 1 - 9/[100E^2].$$

Solve it and get $E = 0.95$.

b) Using the CLT we get

$$n^{1/2}\frac{\bar{X}-\mu}{\sigma} \stackrel{D}{=} Z \sim N(0,1).$$

From this we immediately get the classical formula

$$M = \frac{z_{1-\alpha/2}\sigma}{n^{1/2}},$$

where $z_\beta$ is the beta-quantile, that is $P(Z \le z_\beta) = \beta$.

Plug-in numbers and get

$$M = \frac{(1.645)\sqrt{9}}{\sqrt{100}} = 0.493.$$

Conclusion: As it was expected, Chebyshev's approach is more conservative, but on the other hand, it makes no assumption (it is robust).