Additional Important Facts Involving Random Variables

Transformations

(Discrete Case) For a transformation of a discrete random variable, write out the probability distribution table.

(Continuous Case) For a transformation of a given continuous random variable, use the cdf of the given random variable.
Special Continuous Case:
If the transformation $Y = g(X)$ is 1-1 (i.e. satisfies the horizontal line test) then first find the inverse $h(X) = g^{-1}(X)$. Then the pdf of Y is

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

Sums of random variables

If Y is a sum of random variables, we have seen before how to calculate the expectation and variance of Y. Generally, the expectation of a sum of random variables is the sum of the expectations of each random variable in the sum. The same is not true for variance, *except when the random variables are mutually independent* (i.e. $Cov(X_i, X_j) = 0$ when $i \neq j$.)

Mutually Independent Sum of Random Variables (variance and mgf)

If $X_1, X_2, \cdots, X_n$ are mutually independent random variables and $Y = \sum_{i=1}^{n} X_i$ then

$$Var(Y) = \sum_{i=1}^{n} Var(X_i)$$

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

Note that these last two formulas are true only if the random variables in the sum are mutually independent.

Random Sample of Size *n*

The collection $X_1, X_2, \cdots, X_n$ is a random sample from the distribution *X* means $X_1, X_2, \cdots, X_n$ are mutually independent random variables, each with the same distribution as *X*. Then, for $Y = \sum_{i=1}^{n} X_i$ we have

$$Var(Y) = \sum_{i=1}^{n} Var(X_i) = n \cdot Var(X)$$

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = [M_X(t)]^n$$

Note that these last two formulas are true only if the random variables in the sum are mutually independent.

Certain Independent Sums $Y = \sum_{i=1}^{n} X_i$

$$X_i \sim B(1, p) \Rightarrow Y \sim B(n, p)$$
$$X_i \sim B(n_i, p) \Rightarrow Y \sim B(\sum n_i, p)$$
$$X_i \sim G(p) \Rightarrow Y \sim NB(n, p)$$
$$X_i \sim NB(n_i, p) \Rightarrow Y \sim NB(\sum n_i, p)$$
$$X_i \sim P(\lambda) \Rightarrow Y \sim P(n \cdot \lambda)$$
$$X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow Y \sim N(\sum \mu_i, \sum \sigma_i^2)$$

Note that these formulas are true only if the random variables in the sum are mutually independent.

Central Limit Theorem – Suppose the collection $X_1, X_2, \cdots, X_n$ is a random sample from the distribution *X*. Let $Y_n = \sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$. Then for large *n* (*n* > 30), the distribution of Y_n is approximately normal with mean $E[Y_n] = n \cdot E[X]$ and $Var(Y_n) = n \cdot Var(X)$.

Relationship between Exponential and Poisson Distribution – If the time between events (in some unit of time, say minutes) follows an exponential distribution with mean θ, then the number of claims per minute will follow a Poisson distribution with mean $1/\theta$. In symbols,

$$T \sim EX(mean = \theta) \Rightarrow N \sim P(1/\theta)$$

Maximums and Minimums – (illustrated with 3 independent variables, but can be generalized) Suppose $U = \max(X_1, X_2, X_3)$ and $V = \min(X_1, X_2, X_3)$. In order to get the pdf of *U*, first find the cdf of *U* and take the derivative to get the pdf. In order to get the pdf of *V*, first find the sf, and take the negative of the derivative to get the pdf. Get the cdf of *U* and sf of *V* as follows:

$$F_U(u) = \Pr(U \le u) = \Pr(X_1 \le u) \cdot \Pr(X_2 \le u) \cdot \Pr(X_3 \le u) = F_{X_1}(u) \cdot F_{X_2}(u) \cdot F_{X_3}(u)$$
$$S_V(v) = \Pr(V > v) = \Pr(X_1 > v) \cdot \Pr(X_2 > v) \cdot \Pr(X_3 > v) = S_{X_1}(v) \cdot S_{X_2}(v) \cdot S_{X_3}(v)$$

Mixtures of Distributions – The random variable *W* is a mixture of random variables *X* and *Y* means the pdf of *W* is $f_W(w) = a \cdot f_X(w) + (1-a) \cdot f_Y(w)$. Then

$$\begin{aligned}
f_W(w) &= a \cdot f_X(w) + (1-a) \cdot f_Y(w) \\
F_W(w) &= a \cdot F_X(w) + (1-a) \cdot F_Y(w) \\
s_W(w) &= a \cdot s_X(w) + (1-a) \cdot s_Y(w) \\
M_W(t) &= a \cdot M_X(y) + (1-a) \cdot M_Y(t) \\
E[W^n] &= a \cdot E[X^n] + (1-a) \cdot E[Y^n]
\end{aligned}$$

*****WARNINGS*****

1. It is NOT true that $W = a \cdot X + (1-a) \cdot Y$.
2. A very common mistake is to try write an analogous formula for the variance of *W*. THERE IS NO SUCH FORMULA. In order to find the variance of *W* we use the formula

$$Var(W) = E[W^2] - (E[W])^2$$

Use the above formula for $E[W^n]$ to find $E[W^2]$ and $E[W]$.