1 Monday’s, 10/29/12, notes: Normal Distribution

Note: I am putting the notes for both Monday and Wednesday’s classes here. I am putting the examples for both days under Wednesday’s class.

One drawback of the Normal Distribution is that its cdf is not a simple algebraic formula. There is no closed form solution to the cdf of a Normal. Therefore, in order to find any probability associated with a Normal($\mu$, $\sigma^2$) random variable we need to do an algebraic trick that is called standardizing a Normal r.v. To understand this concept, first we need to introduce the variable Z. In Statistics, Z is reserved for a Normal($\mu = 0, \sigma = 1$) random variable. Z is referred to as the Standard Normal. Our ”trick” is to turn a Normal($\mu$, $\sigma^2$) into a Normal($\mu = 0, \sigma = 1$) random variable. This is done by the following formula:

$$Z = \frac{X - \mu}{\sigma}.$$ 

Unlike other continuous random variables, the pdf and cdf for Z are not labeled with f and F. Instead, they are labeled with $\phi$ and $\Phi$ respectively. Because of the importance of Z in Statistics, it gets its own letter to represent its pdf and cdf. However, since Z is a Normal r.v. its cdf does not exist in closed form either. Instead, we have a table of probabilities. The one we will use in this course is on the course web site as ”Normal Table”. Please print this pdf off and bring it with you to every class.

If X is a Normal($\mu$, $\sigma^2$) r.v., then $P(c < X < d) = \Phi(\frac{d - \mu}{\sigma}) - \Phi(\frac{c - \mu}{\sigma})$. In other words, we can relate the cdf of X to the cdf of Z. $F_X(x) = \Phi(\frac{x - \mu}{\sigma})$. Recall that a Normal r.v. is symmetric. This actually implies the following: $\Phi(-z) = 1 - \Phi(z)$. This is useful for $P(Z \geq z) = 1 - \Phi(z) = \Phi(-z)$.

Now that we can calculate probabilities for a Normal r.v., there are 2 other main topics to discuss. The first is about sums of independent Normal random variables. Let $X_i$ denote mutually independent Normal random variables with parameters $\mu_i$ and $\sigma_i$ respectively. Their sum has mean equal to the sum of the $\mu_i$ and variance equal to the sum of the variances. If we let $Y = \sum_{i=1}^{n} X_i$ then $Y \sim \text{Normal}(\mu_Y = \sum_{i=1}^{n} \mu_i, \sigma_Y^2 = \sum_{i=1}^{n} \sigma_i^2)$. This can be applied to any number of Normal random variables (provided that they are mutually independent). (Quick aside: This provides motivation for the CLT, which a lot of you will see in MGMT 305.)

The last topic is that we have now our third and final approximation. We can use a Normal Distribution to approximate a Binomial Distribution if n is large and p is moderate (close to .5). Our rule of thumb for this approximation to be valid is that both $np > 5$ and $n(1-p) > 5$. If $X \sim \text{Binomial}(n,p)$ and the approximation holds, then the approximation, $X^* \sim N(\mu = np, \sigma^2 = np(1-p))$. One caveat to this approximation is that we are approximating a discrete distribution
(the Binomial) with a continuous distribution (the Normal). One thing that we know about these
types of distributions is that discrete r.v.s have point probabilities, but continuous r.v.s do not.
In order to account for this, we use the continuity correction. This involves either adding or
subtracting a half from the x value accordingly.

2  Wednesday’s, 10/31/12, notes: Normal Approximation to the
Binomial

Example 11.1 Let us examine Z. Find the following probabilities with respect to Z: at most
-1.75, at most 1.75, between -2 and 2 inclusive, less than .5. Find the following with respect to
Z: the value such that 20.3% are higher than it, the 4.65\textsuperscript{th} percentile, and the values representing
the middle 96.6% of the distribution.

Example 11.2 Let X be Normal with a mean of 20 and a variance of 49. Find the following
probabilities: X is between 15 and 23; X is more than 12 knowing it is less than 20; given X is
less than 28, the probability that it is more than 16; and that it is more than 31. What is the
value that is smaller than 20% of the distribution?

Example 11.3 Let $X_1$, $X_2$, and $X_3$ be mutually independent, Normal random variables. Let
their means and standard deviations be 3k and k for k = 1, 2, and 3 respectively. Find the fol-
lowing distributions: $\sum_{i=1}^{3} X_i$, $X_1 + X_2 - X_3$, $2X_1 - 3X_3 + 4X_3$. Call the previous distributions
S, T, and V respectively. Find the following percentiles for S, T, and V respectively, 83\textsuperscript{rd}, 63\textsuperscript{rd},
and 42\textsuperscript{nd}. Find the following probabilities: S is bigger than V’s mean, T is smaller than half of
S’s variance, and V is bigger than T’s 99\textsuperscript{th} percentile.

Example 11.4 Suppose a class has 400 students (to begin with), that each student drops in-
dependently of any other student with a probability of .07. Let X be the number of students
that finish this course. Find the probability that X is between 370 and 373 inclusive? Is an
approximation appropriate for the number of students that finish the course? If so, what is this
distribution and what are the value(s) of its parameter(s)? For the following probabilities, if an
approximation is appropriate, use the approximation; otherwise, use the exact distribution. Find
the probability that is between 370 and 373 inclusive, that X is at least 375, that X is at most
370, that X is between 360 and 380, and that X is between 360 and 380 inclusive.

Example 11.5 SAT Math scores follow a Normal distribution with a mean of 533 and a standard
deviation of 116. Assuming that scores above 800 get truncated to 800, what percent of scores
were reported as 800? The middle 50\% of SAT Math scores at Purdue in 2011 were reported as
550 to 690. What percent of all SAT Math scores were in this range? Notre Dame’s middle 50\% are
between 680 and 770. What percent of all scores are below Notre Dame’s 75\textsuperscript{th} percentile?
What percent of all scores are above Notre Dame’s 25\textsuperscript{th} percentile?

Example 11.6 Colin and Mike are wasting their childhood playing ping pong in Colins base-
ment. Since they have spent so much time in the basement playing ping pong, pool, and darts, they are famished. They decide to order Chinese food with extra teriyaki sauce for delivery. If the food will arrive according to a normal distribution with mean of 20 minutes and standard deviation of 5 minutes, what is the probability that the two kids have to wait more than 32 minutes for their food? What is the probability that they wait less than 15 minutes? What is the probability that they wait less than 26 minutes, knowing that they wait at least 12 minutes?

**Example 11.7** Suppose you and 4 of your best friends are migrating west. You are the local physician. Suppose you decide to hunt buffalo. On average buffalo have 800 lbs. of edible meat with a standard deviation of 75 lbs. If your party comes back to the trail with one buffalo, what is the probability that you come back with less than 700 lbs. of edible meat? If you need 925 pounds of edible meat to make it all the way to Independence, Missouri, what is the probability that your 1 buffalo will last you until Independence, Missouri? What amount of edible meat is less than 29% of the distribution?

**Example 11.8** Wish by NIN is a 3 minute and 36 second long song. Suppose pyrotechnics on average last for 2 minutes, but they have a standard deviation of 53 seconds. Suppose NIN use pyrotechnics at the beginning of "Wish". What is the probability that the fog will still mask Trent Reznor at the end of Wish?

**Example 11.9** A male yeti's height is normally distributed with a mean of 84 inches and a standard deviation of 7 inches. Since, yetis seem to elude people, we will not make a question about the probability of a specific yeti, but of yetis in general. What are the 25th, 48th, and 67th percentiles for height of a yeti?

3 Friday’s, 11/2/12, notes: Poisson Process

For a specified event that occurs randomly in continuous time, an important application of probability theory is in modeling the number of times such an event occurs. The following are several examples of such random phenomenon.

- The number of patients that arrive at a hospital emergency room.
- The number of customers that enter a particular bank.
- The number of accidents at an intersection.
- The number of alpha particles emitted by a radioactive substance.

Consider an event that occurs randomly and homogenously in continuous time at an average rate of \( \lambda \) per unit of time. We will refer to the occurrence of the event as a success. If we begin counting successes at time 0, and, for each time, \( t \geq 0 \), we let \( N(t) \) = the number of successes by time \( t \) (\( \leq t \)). Automatically, this implies that \( N(0) = 0 \). We say such a counting process is a Poisson process with rate \( \lambda \) if 2 more properties hold. Namely, if:
• \( N(t) \): \( t \geq 0 \) has independent increments (as long as the two time intervals have no overlap, they are independent).

• \( N(t) - N(s) \), which is the number of successes in the time interval \((s, t]\), is distributed as Poisson\((\lambda(t-s))\) for \(0 \leq s < t < \infty\).

As indicated by previous examples, the Poisson Process can be used to model arrivals. It is also used for waiting times and interarrival times.

For each \( n \in \mathbb{N} \), we let \( W_n \) denote the time of the occurrence of the \( n^{th} \) event. That is the time at which the \( n^{th} \) success occurs. If \( W_3 = 10.34 \), that means the 3\(^{rd} \) success occurred at a time of 10.34. The random variable \( W_n \) is called the \( n^{th} \) waiting time. The elapsed time between the occurrence of the \((n - 1)^{st}\) and \( n^{th} \) events is denoted by \( I_n \) and is called the \( n^{th} \) interarrival time.

So, we have the following 2 relationships:

\[
W_n = \sum_{j=1}^{n} I_j
\]

\[
I_n = W_n - W_{n-1}
\]

One nice property of a Poisson Process with rate \( \lambda \) is that the interarrival times, or \( I_n \)'s are iid Exponential random variables with rate parameter \( \lambda \).

There is one more property of a Poisson Process that is quite useful. Suppose we have \( W_t = n \). This means that we had \( n \) successes on the interval \([0, t]\). These successes are independent Uniform\((0, t)\) random variables. Keep in mind that time increments are independent for a Poisson random variable if there is no overlap. Knowing \( W_t = n \), if we looked at the distribution of the number of successes on the interval \([0, t]\), how would these be distributed?

**Example 11.10** Suppose that phone calls arrive at a switchboard according to a Poisson Process at a rate of 2 per minute. Let \( X \) be the number of calls between 9:30 and 9:45. Find the distribution of \( X \). Let \( T \) be the time between the 8\(^{th}\) and 9\(^{th}\) calls. What is the distribution of \( T \)? What is the probability that exactly 10 calls (total) come in the next 4 minutes? What is the probability that the next call comes in 30 seconds and the second call comes at least 45 seconds after that? Given there are exactly 7 calls in 3 minutes, what is the probability that they all came in the last minute?

**Example 11.11** Each time a student logs on to their ITaP account, the computer sends a request for the student’s profile to the main ITaP database. Suppose that these profile requests come to the main database according to a Poisson Process at a rate of 9 per minute. What is the probability that between 8 and 11 (inclusive) profile requests go to ITaP in a given minute? On average, how many profile requests arrive in an hour period? What is the probability of 7 profile requests in a 1-minute interval followed by 19 profile requests in the subsequent 2-minute interval? How long, on average, does it take between successive profile requests? What is the probability that the next profile request takes more than 15 seconds? What is the probability that the next profile request takes at most 22 seconds? It we know that 13 profile requests occurred between 12:00:00
AM and 12:01:30 AM, what is the probability that 5 profile requests occurred between 12:00:50 and 12:01:20?

**Example 11.12** Customers arrive at Scotty’s at a rate of .5 per minute. (Assume all customers arrive independently of all other customers.) What is the probability that 10 customers arrive in the next 15 minutes? What is the probability that 10 customers arrive in each of the next 4 15-minute intervals? How long on average does it take for the next customer to arrive? What is the probability that $I_1$ is more than 20 seconds, $I_2$ is more than 30 seconds, and $I_3$ is less than 15 seconds?