14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 14

Konrad Menzel

April 2, 2009

1 Conditional Expectations

Example 1 *Each year, a firm's R&D department produces X innovations according to some random process, where $\mathbb{E}[X] = 2$ and $\text{Var}(X) = 2$. Each invention is a commercial success with probability $p = 0.2$ (assume independence). The number of commercial successes in a given year are denoted by S. Since we know that the mean of $S \sim B(x, p) = xp$, conditional on $X = x$ innovations in a given year, xp of them should be successful on average.*

The *conditional* expectation of X given Y is the expectation of X taken over the conditional p.d.f.:
Definition 1

$$\mathbb{E}[Y|X] = \begin{cases} \sum_y y f_{Y|X}(y|X) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|X)dy & \text{if } Y \text{ is continuous} \end{cases}$$

Note that since $f_{Y|X}(y|X)$ carries the *random variable* X as its argument, the conditional expectation is also a random variable. However, we can also define the conditional expectation of Y given a particular value of X,

$$\mathbb{E}[Y|X = x] = \begin{cases} \sum_y y f_{Y|X}(y|x) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy & \text{if } Y \text{ is continuous} \end{cases}$$

which is just a number for any given value of x as long as the conditional density is defined.
Since the calculation goes exactly like before, only that we now integrate over the *conditional* distribution, won't do a numerical example (for the problem set, just apply definition). Instead let's discuss more qualitative examples to illustrate the difference between conditional and unconditional examples:

Example 2 ***(The Market for "Lemons")*** *The following is a simplified version of a famous model for the market for used cars by the economist George Akerlof. Suppose that there are three types X of used cars: cars in an excellent state ("melons"), average-quality cars ("average" not in a strict, statistical, sense), and cars in a poor condition ("lemons"). Each type of car is equally frequent, i.e.*

$$P(\text{"}lemon\text{"}) = P(\text{"}average\text{"}) = P(\text{"}melon\text{"}) = \frac{1}{3}$$

The seller and a buyer have the following (dollar) valuations Y_S and Y_B, respectively, for each type of cars:

Type	*Seller*	*Buyer*
"Lemon"	*5,000$*	*6,000$*
"Average"	*6,000$*	*10,000$*
"Melon"	*10,000$*	*11,000$*

The first thing to notice is that for every type of car, the buyer's valuation is higher than the seller's, so for each type of car, trade should take place at a price between the buyer's and the seller's valuations. However, for used cars, quality is typically not evident at first sight, so if neither the seller nor the buyer know the type X of a car in question, their expected valuations are, by the law of iterated expectations

$$\begin{aligned}
\mathbb{E}[Y_S] &= \mathbb{E}[Y_S|\text{"}lemon\text{"}]P(\text{"}lemon\text{"}) + \mathbb{E}[Y_S|\text{"}average\text{"}]P(\text{"}average\text{"}) + \mathbb{E}[Y_S|\text{"}melon\text{"}]P(\text{"}melon\text{"}) \\
&= \frac{1}{3}(5,000 + 6,000 + 10,000) = 7,000 \\
\mathbb{E}[Y_S] &= \mathbb{E}[Y_B|\text{"}lemon\text{"}]P(\text{"}lemon\text{"}) + \mathbb{E}[Y_B|\text{"}average\text{"}]P(\text{"}average\text{"}) + \mathbb{E}[Y_B|\text{"}melon\text{"}]P(\text{"}melon\text{"}) \\
&= \frac{1}{3}(6,000 + 10,000 + 11,000) = 9,000
\end{aligned}$$

so trade should still take place.
But in a more realistic setting, the seller of the used car knows more about its quality than the buyer (e.g. history of repairs, accidents etc.) and states a price at which he is willing to sell the car. If the seller can perfectly distinguish the three types of cars, whereas the buyer can't, the buyer should form expectations conditional *on the seller willing to sell at the quoted price.*
If the seller states a price less than $6,000$ *dollars, the buyer knows for sure that the car is a "lemon" because otherwise the seller would demand at least* $6,000$*, i.e.*

$$\mathbb{E}[Y_B|Y_S < 6000] = \mathbb{E}[Y_B|\text{"}lemon\text{"}] = 6000$$

and trade would take place. However, if the car was in fact a "melon", the seller would demand at least $10,000$ *dollars, whereas the buyer would pay at most*

$$\mathbb{E}[Y_B|Y_S \leq 10,000] = \mathbb{E}[Y_B] = 9,000 < 10,000$$

so that the seller won't be able to sell the high-quality car at a reasonable price.
The reason why the market for "melons" breaks down is that in this model, the seller can't credibly assure the buyer that the car in question is not of lower quality, so that the buyer factors the possibility of getting the bad deal into his calculation.

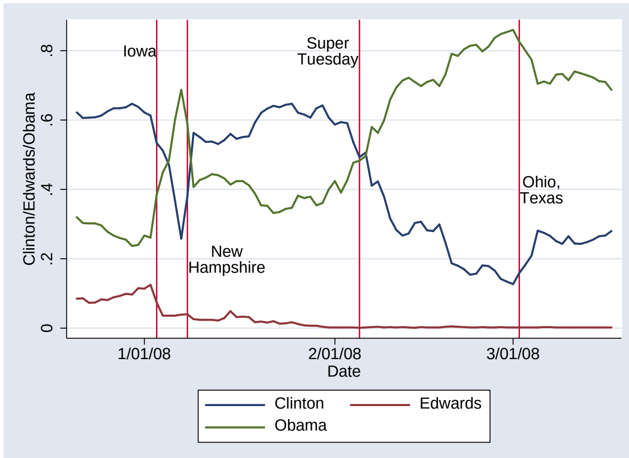
Example 3 *In this example, we are going to look at data on the 2008 presidential nominations from the IEM Political Markets, an internet platform in which people bet on future political events (the data can be obtained from*
http://www.biz.uiowa.edu/iem/markets/data_nomination08.html).
The market works as follows: for each political candidate i, participants can buy a contract which pays

$$Y_i = \begin{cases} 1 \text{ dollar} & \text{if candidate } i \text{ wins nomination} \\ 0 \text{ dollars} & \text{otherwise} \end{cases}$$

At a given point in time t, participants in the market have additional outside information, which we'll call X_t, as e.g. the number of delegates won so far, the "momentum" of a candidate's campaign, or statements by staff about the candidate's campaign strategy.
Given that additional information, the expected value of the contract is

$$\mathbb{E}[Y_i|X_t = x] = \sum_{y_i} y_i f_{Y_i|X_t}(y_i|x) = 1 \cdot P(Y_i = 1|X_t = x) + 0 \cdot P(Y_i = 0|X_t = x) = P(Y_i|X_t)$$

In other words, the dollar amount traders should be willing to pay for the contract for candidate i equals the probability that i wins his/her party's nomination given the available information *at time t.*
Let's look at the prices of contracts for the main candidates of the Democratic party over the last 3 months: I put three vertical lines for 3 events which revealed important information about the candidates' likelihood of winning the Democratic nomination:



- *the Iowa caucus in which Barack Obama won against Hillary Clinton by a significant margin*
- *the New Hampshire primaries which were seen as Hillary Clinton's "comeback" after the defeat in Iowa*
- *the primaries in Ohio and Texas, two major states which were won by Hillary Clinton*

We can see steep changes in the conditional expectations after each of these events, illustrating how the market updates its "beliefs" about the candidates' chances of securing their parties nomination.
In Financial Economics, this type of contracts which pays 1 dollar if a particular state of nature is realized are also called Arrow-Debreu securities.

An important relationship between conditional and unconditional expectation is the Law of Iterated Expectations (a close "cousin" of the Law of Total Probability which we saw earlier in the lecture):

Proposition 1 ***(Law of Iterated Expectations)***

$$\mathbb{E}\left[\mathbb{E}[Y|X]\right] = \mathbb{E}[Y]$$

Proof: Let $g(x) = \mathbb{E}[Y|X = x]$, which is a function of x. We can now calculate the expectation

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x)dx = \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_X(x)dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y \frac{f_{XY}(x,y)}{f_X(x)} dy \right) f_X(x)dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x,y)dydx \\
&= \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{XY}(x,y)dx \right) dy \\
&= \int_{-\infty}^{\infty} y f_Y(y)dy = \mathbb{E}[Y]
\end{aligned}$$

Proposition 2 ***(Conditional Variance / Law of Total Variance)***

$$\mathrm{Var}(Y) = \mathrm{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\mathrm{Var}(Y|X)]$$

This result is also known as the ANOVA identity, where ANOVA stands for Analysis of Variance. In particular, since $\mathrm{Var}(Y|X) \geq 0$, it follows that

$$\mathrm{Var}(Y) \geq \mathbb{E}[\mathrm{Var}(Y|X)]$$

which can, loosely speaking, be read as "knowing X decreases the variance of Y."

PROOF: We can rewrite

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\mathrm{Var}(Y|X)] &= \left(\mathbb{E}[\mathbb{E}[Y|X]^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2\right) + \left(\mathbb{E}\left[\mathbb{E}[Y^2|X]\right] - \mathbb{E}\left[\mathbb{E}[Y|X]^2\right]\right) \\
&= \mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2 + \mathbb{E}[Y^2] - \mathbb{E}[\mathbb{E}[Y|X]^2] \\
&= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \mathrm{Var}(Y)
\end{aligned}$$

where the first equality uses the property $\mathrm{Var}X = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, the second step uses the law of iterated expectations, and in the last step the first and last term cancel, which completes the argument $\square$

Example 4 *Each year, a firm's R&D department produces X innovations according to some random process, where $\mathbb{E}[X] = 2$ and $\mathrm{Var}(X) = 2$. Each invention is a commercial success with probability $p = 0.2$ (assume independence). The number of commercial successes in a given year are denoted by S.*

(a) Suppose we have 5 new innovations this year. What is the probability that S of them are commercial successes? - the conditional *p.d.f of S given $X = 5$ is that of a binomial, so e.g.*

$$P(S = 2|X = x) = \begin{pmatrix} 5 \\ 3 \end{pmatrix} (0.2)^3(1 - 0.2)^2 \approx 0.05$$

(b) Given 5 innovations, what is the expected number of successes? - since $S|X = 5 \sim B(5, 0.2)$, we can use the result from last class

$$\mathbb{E}[S|X = 5] = \mathbb{E}[B(5, 0.2)] = 0.2 \cdot 5 = 1$$

(c) What is the unconditional *expected value of innovations? - By the law of iterated expectations,*

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|X]] = \mathbb{E}[pX] = 0.2\mathbb{E}[X] = 0.2 \cdot 2 = 0.4$$

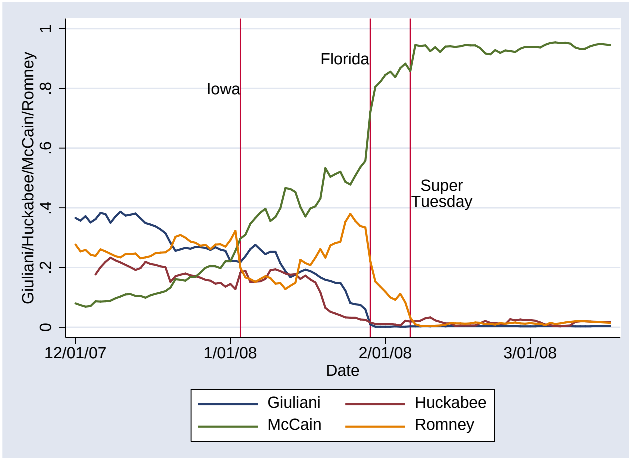
since we assumed $\mathbb{E}[X] = 2$.

(d) What is the unconditional *variance of S? - Recall the law of total variance:*

$$\begin{aligned}
\mathrm{Var}(S) &= \mathrm{Var}(\mathbb{E}[S|X]) + \mathbb{E}[\mathrm{Var}(S|X)] \\
&= \mathrm{Var}(Xp) + \mathbb{E}[p(1-p)X] = p^2\mathrm{Var}(X) + p(1-p)\mathbb{E}[X] = 0.04 \cdot 2 + 0.16 \cdot 2 = 0.4
\end{aligned}$$

This is an example of a mixture *of binomials, i.e. conditional on X, we have a binomial distribution for S. We can then use the law of iterated expectations to obtain the total number of successes.*

Example 5 *(IEM Political Markets, continued) If we look at the Republican primaries last year, much of the uncertainty already was resolved by Super Tuesday. Say, the conditioning variable X_t is the number*



of pledged delegates by date t, we can compare the "unconditional" means before the Iowa primaries to the "conditional" means after Super Tuesday in light of the law of total variance,

$$\mathrm{Var}(Y_i) = \mathbb{E}[\mathrm{Var}(Y_i|X_t)] + \mathrm{Var}(\mathbb{E}[Y_i|X_t])$$

We can see that while before the Iowa elections, the $\mathbb{E}[Y_i]$ for the main candidates were in an intermediate range from 10% to 40% with lots of variation. However, after Super Tuesday, prices (i.e. $\mathbb{E}[Y_i|X_t]$) moved close to 0 or 1, and the "wiggles" have become really tiny. So, in terms of the conditional variance formula, the largest part of the ex ante variance $\mathrm{Var}(Y_i)$ was uncertainty about the conditional mean after Super Tuesday, $\mathrm{Var}(\mathbb{E}[Y_i|X_t])$, whereas the contribution of the conditional variance $\mathrm{Var}(Y_i|X_t)$ seems to be relatively small.
If we compare this to the graph for the Democratic race, for the latter, there is still a lot of movement after Super Tuesday, so that conditioning on the number of pledged delegates X_t by Super Tuesday doesn't take out much of the variance, i.e. $\mathrm{Var}(Y_i|X_t)$ is still considerably large. As an aside, an often cited reason why the Republican race was decided so much earlier is that in the Republican primaries, in each state, delegates are allocated not proportionally to each candidate's vote share (which is the rule for most primaries of the Democratic party), but the winner-takes-all rule, so that even very close victories in the first primaries can get a candidate far enough ahead to make it extremely hard for competitors to catch up.

2 Special Distributions

In the lecture, we already saw three commonly used distributions, the binomial, the uniform and the exponential. Over the next two lectures, we are going to expand this list by a few more important examples, and we'll start with the most frequently used of all, the normal distribution.

2.1 Recap: Distributions we already saw in Class

Definition 2 *X is binomial distributed with parameters (n, p), $X \sim B(n, p)$ if its p.d.f. is given by*

$$f_X(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x} & \text{if } x \in \{0, 1, \dots, n\} \\ 0 & \text{otherwise} \end{cases}$$

We showed that for $X \sim B(n,p)$,

$$\begin{aligned} \mathbb{E}[X] &= np \\ \mathrm{Var}(X) &= np(1-p) \end{aligned}$$

Definition 3 *X is uniformly distributed over the interval $[a,b]$, $X \sim U[a,b]$, if it has p.d.f.*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Definition 4 *X is exponentially distributed with parameter λ, $X \sim E(\lambda)$, if it has p.d.f.*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

2.2 Standardized Random Variable

Sometimes, it is useful to look at the following *standardization* Z of a random variable X

$$Z := \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}}$$

Using the rules for expectations and variances derived in the last couple of lectures,

$$\mathbb{E}[Z] = \frac{\mathbb{E}[X - \mathbb{E}[X]]}{\sqrt{\mathrm{Var}(X)}} = 0$$

and

$$\mathrm{Var}(X) = \frac{\mathrm{Var}(X - \mathbb{E}[X])}{\mathrm{Var}(X)} = \frac{\mathrm{Var}(X)}{\mathrm{Var}(X)} = 1$$

If we normalize random variables in this way, it's easier to compare shapes of different distributions independent of their scale and location.

2.3 The Normal Distribution

The normal distribution corresponds to a continuous random variable, and it turns out that it gives good approximations to a large number of statistical experiments (we'll see one example in a second, more on this when we discuss the Central Limit Theorem next week).

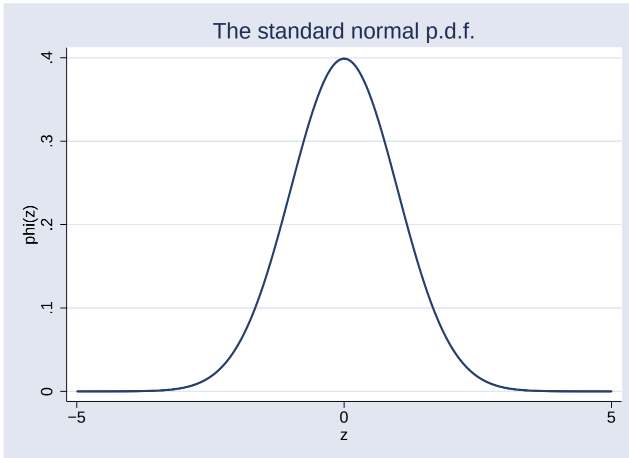
Definition 5 *A random variable Z follows a* standard normal *distribution - in symbols $Z \sim N(0,1)$ - if its p.d.f. is given by*

$$f_Z(z) = \varphi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

for any $z \in \mathbb{R}$. Its c.d.f. is denoted by

$$\Phi(z) := P(Z \leq z) = \int_{-\infty}^{z} \varphi(s)ds$$

The c.d.f. of a standard normal random variable doesn't have a closed-form expression (but can look up values in tables, also programmed in any statistical software package). The p.d.f. $\varphi(z)$ has a characteristic bell shape and is symmetric around zero:



2.3.1 Important Properties of the Standard Normal Distribution

Property 1 *For a standard normal random variable Z,*

$$\mathbb{E}[Z] = 0$$

and

$$\mathrm{Var}(Z) = 1$$

As a first illustration why the normal distribution is very useful, it turns out that Binomial random variables are approximated by the normal for a large number n of trials:

Theorem 1 ***(DeMoivre-Laplace Theorem)*** *If $X \sim B(n, p)$ is a binomial random variable, then for any numbers $c \leq d$,*

$$\lim_{n\to\infty} P\left(c \leq \frac{X - np}{\sqrt{np(1-p)}} < d\right) = \lim_{n\to\infty} P\left(c < \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}} \leq d\right) = \int_c^d \varphi_Z(z)dz$$

Notice that the transformation of the binomial variable to

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}}$$

is in fact a standardization. This result says that for large n, the probability that the *standardized* binomial random variable X falls inside the interval $(c, d]$ is approximately the same as for a standard normal random variable. As an illustration, I plotted the binomial p.d.f. for increasing values of n, and then applied the normalization.

For $n = 50$, the shape of the bar graph looks already relatively similar to the bell-shape of the normal density. Note that in particular the skewness of the distribution for small n (due to the small "success" probability $p = \frac{1}{4}$) washes out almost entirely.

Example 6 *In order to see why this type of approximation is in fact useful for practical purposes, say $p = \frac{1}{5}$, and we want to calculate the probability that in a sequence of n trials, we have at least 25%*

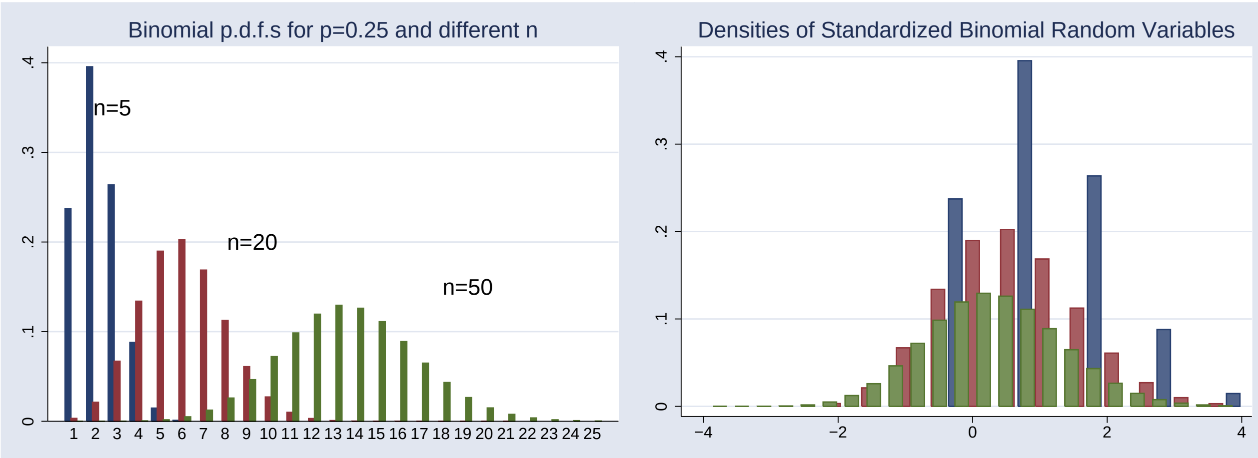


Figure 1: Illustration of the DeMoivre-Laplace Theorem

successes.
If $n = 5$, the probability of having no more than 25% successes can be calculated as

$$P(X \leq 1.25) = P(X = 0) + P(X = 1) = \binom{5}{0}(1-p)^5 + \binom{5}{1}p(1-p)^4 = \frac{4^5}{5^5} + 5\frac{4^4}{5^5} = \frac{1280}{3125} \approx 40.96\%$$

What if $\tilde{X} \sim B\left(100, \frac{1}{5}\right)$, i.e. we increase n to 100? In principle, we could calculate

$$P(\tilde{X} \leq 25) = P(X = 0) + P(X = 1) + \ldots + P(X = 25)$$

So we'd have to calculate each summand separately, and since there are pretty many of those, this will be very cumbersome. Alternatively, we could limit ourselves to a good approximation using the DeMoivre-Laplace Theorem. The standardized version of $\tilde{X}$ is given by

$$Z = \frac{\tilde{X} - 100 \cdot \frac{1}{5}}{\sqrt{100 \cdot \frac{1}{5} \cdot \frac{4}{5}}} = \frac{\tilde{X} - 20}{4}$$

Therefore,

$$P(\tilde{X} \leq 25) = P(\tilde{X} - 20 \leq 5) = P\left(Z \leq \frac{5}{4}\right) \approx \int_{-\infty}^{\frac{5}{4}} \phi(z)dz \approx 89.44\%$$

How good is this approximation? I did the calculation, and I obtained $P(\tilde{X} \leq 25) \approx 91.25\%$. If we repeat the same exercise for $n = 200$, I obtain the "exact" binomial probability $P(\bar{X} \leq 50) \approx 96.55\%$ and the normal approximation $P(\bar{X} \leq 50) \approx 96.15\%$.

For $Z \sim N(0, 1)$, we also call any random variable

$$X = \mu + \sigma Z$$

a *normal* random variable with mean μ and variance σ^2, or in symbols

$$X \sim N(\mu, \sigma^2)$$

What is the p.d.f. of X? By the change of variables formula we saw earlier in this class,

$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

We can extend the same argument by noting that the linear transformation of a normal random variable X_1 is again normal.

Proposition 3 *If $X_1 \sim N(\mu, \sigma^2)$ and $X_2 = a + bX_1$, then*

$$X_2 \sim N(a + b\mu, b^2\sigma^2)$$

You can check this again using the change of variables formula.
We already saw that the expectation of the sum of n variables $X_1, \ldots, X_n$ is the sum of their expectations, and that the variance of n *independent* random variables is the sum of their variances. If the X_i's are also normal, then their sum is as well:

Proposition 4 *If $X_1, \ldots, X_n$ are independent normal random variables with $X_i \sim N(\mu_i, \sigma_i^2)$, then*

$$Y = \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

While in the general, we'd have to go through the convolution formula we saw a few weeks ago, for the sum of normals, we therefore only have to compute the expectation and variance of the sum, and know the p.d.f. of the sum right away:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sum\sigma_i^2}}e^{-\frac{(y-\sum\mu_j)^2}{2\sum\sigma_i^2}}$$

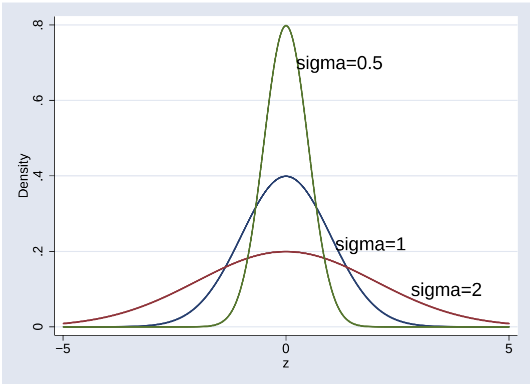


Figure 2: Normal Density for Different Values of σ

2.3.2 Using Tabulated Values of the Standard Normal

If $X \sim N(\mu, \sigma)$, we can give a rough estimate of the probability with which X is no further than one or two standard deviations away from its mean:

$$\begin{aligned}
P(\mu - 1\sigma \le X \le \mu + 1\sigma) = \Phi(1) - \Phi(-1) &\approx 68\% \\
P(\mu - 2\sigma \le X \le \mu + 2\sigma) = \Phi(2) - \Phi(-2) &\approx 95\% \\
P(\mu - 3\sigma \le X \le \mu + 3\sigma) = \Phi(3) - \Phi(-3) &\approx 99.7\%
\end{aligned}$$

i.e. most of the mass of the distribution is within one or two standard deviations of the mean. It's useful to remember these three quantities in order to get rough estimates of normal probabilities if you don't have a tabulation of the c.d.f. at hand.

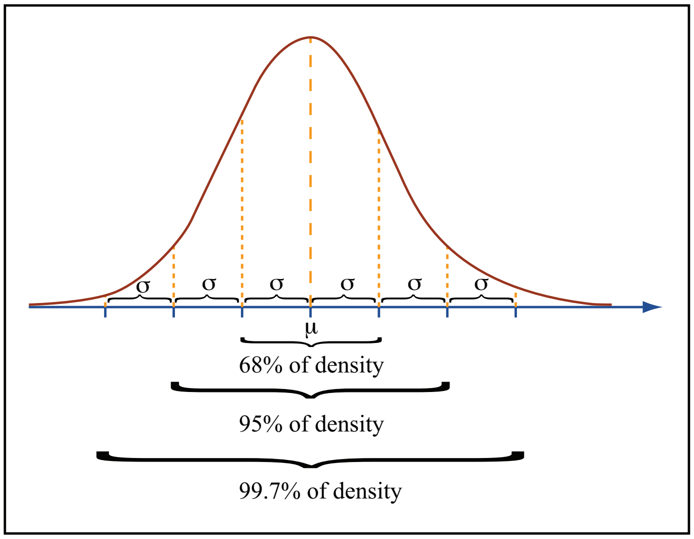


Image by MIT OpenCourseWare.

Since the standard normal distribution is so commonly used, you'll find tabulated values of the c.d.f. $\Phi(z)$ in any statistics text.
Often those tables only contain values $z \leq 0$, but you can obtain the c.d.f. at a value $\tilde{z} > 0$ using symmetry of the distribution:

$$\Phi(\tilde{z}) = 1 - \Phi(-\tilde{z})$$

E.g. if we want to know $P(Z \leq 1.95)$, we can look up $P(Z \leq -1.95) = 0.0256$, and calculate $P(Z \leq 1.95) = 1 - P(Z \leq -1.95) = 0.9744$.

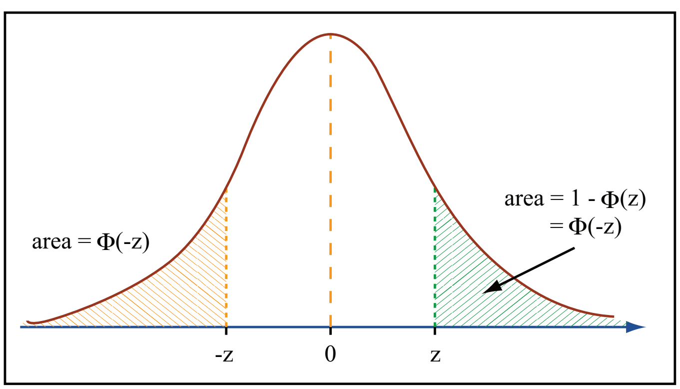


Image by MIT OpenCourseWare.

In general, if $X \sim N(\mu, \sigma)$, we can obtain probabilities of the type $P(a \leq X \leq b)$ for $a \leq b$ using the following steps:

1. standardize the variable, i.e. rewriting the event as

$$P(a \leq X \leq b) = P\left(a \leq \mu + \sigma Z \leq b\right) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

for a *standard* normal random variable Z

2. restate the probability in terms of the standard normal c.d.f., $\Phi(\cdot)$:

$$P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

3. look up the values for the values of the standard normal c.d.f. at the values calculated above in order to obtain the probability.

2.4 Digression: Drawing Standard Normal Random Variables

We already saw that it is possible to convert uniform random draws to any other continuous distribution via the integral transformation (see the lectures on transformations of random variables). What if you don't have a computer? Around 1900, the famous statistician Francis Galton came up with a clever mechanical device for simulating normal random variables using dice:[1]
The three different dice shown in Figure 3 were rolled one after another, while the experimenter fills up a list of random draws in the following manner: the first die gives the actual values (you always read off what is on the bottom of the side of the die facing you), where the stars are first left empty, and later filled up with rolls of the second die. Finally, the third die gives sequences of pluses and minuses which are put in front of the draws put down as we went along with the first two dice.

The numbers on the dice were specifically chosen as evenly spaced percentiles of the positive half of the standard normal distribution in order to ensure that the outcome would in fact resemble a standard normal random variable.

2.5 Functions of Standard Normals: chi-squared, t- and F-distribution

Due to the importance of the standard normal distribution for estimation and testing, some functions of standard normal random variables also play an important role and are frequently tabulated in statistics tests. For now we'll just give definitions for completeness, but we'll get back to applications in the last third of the class. I'm not going to give the corresponding p.d.f.s since they are of limit practical use for us.

Definition 6 *If $Z_1, Z_2, \ldots, Z_k$ are independent with $Z_i \sim N(0,1)$, $Y = \sum_{i=1}^k Z_i^2$ is said to follow a* chi-squared distribution with k degrees of freedom*, in symbols*

$$Y \sim \chi_k^2$$

Here "degrees of freedom" refers to the number of independent draws that are squared and summed up. The expectation of a chi-squared is given by the degrees of freedom,

$$Y \sim \chi_k^2 \Rightarrow \mathbb{E}[Y] = \sum_{i=1}^{k} \mathbb{E}[X_i^2] = k$$

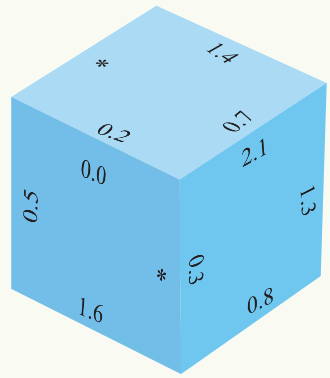
Definition 7 *If $X \sim N(0,1)$ and $Y \sim \chi_k^2$, then*

$$T = \frac{Z}{\sqrt{Y}} \sim t_k$$

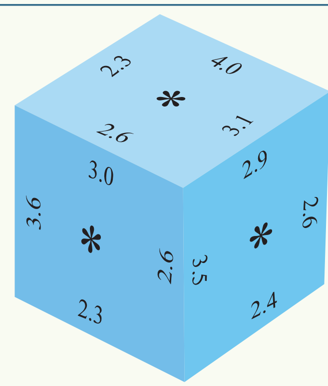
is said to follow the (student) t-distribution with k degrees of freedom*.*

[1] See Stigler, S. (2002): Statistics on the Table

Values for Die I			
0.03	0.51	1.04	1.78
0.11	0.59	1.14	1.95
0.19	0.67	1.25	2.15
0.27	0.76	1.37	(2.40)
0.35	0.85	1.50	(2.75)
0.43	0.94	1.63	(3.60)



Values for Die II			
2.29	2.51	2.77	3.25
2.32	2.55	2.83	3.36
2.35	2.59	2.90	3.49
2.39	2.64	2.98	3.65
2.43	2.68	3.06	4.00
2.47	2.72	3.15	4.55



Values for Die III			
+ + + +	+ − − +	− − + +	+ − +
+ + + −	+ − − −	− − + −	+ − −
+ + − +	− + + +	− − − +	− + +
+ + − −	− + + −	− − − −	− + −
+ − + +	− + − +	+ + +	− − +
+ − + −	− + − −	+ + −	− − −

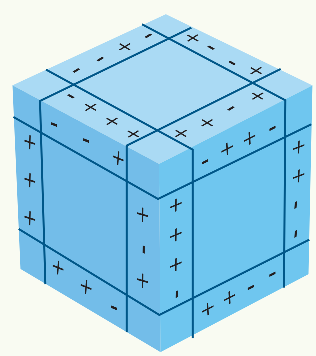


Image by MIT OpenCourseWare.

Figure 3: Three types of Galton's dice. They are 1.25 in. cubes, date from 1890, and are used for simulating normally distributed random numbers. Adapted from Stigler, S. (2002): Statistics on the Table: The History of Statistical Concepts and Methods

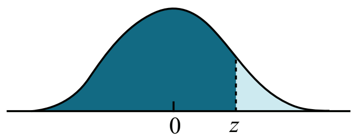
For a large value k for the degrees of freedom, the t-distribution is approximated well by the standard normal distribution.

Definition 8 *If* $Y_1 \sim \chi^2_{k_1}$ *and* $Y_2 \sim \chi^2_{k_2}$*, then*

$$F = \frac{Y_1/k_1}{Y_2/k_2} \sim F(k_1, k_2)$$

is said to follow an F-distribution with (k_1, k_2) degrees of freedom.

Cumulative areas under the standard normal distribution



(Cont.)

z	0	1	2	3	4	5	6	7	8	9
-3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0238	0.0233
-1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0300	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0570	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3112
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Cumulative areas under the standard normal distribution

(Cont.)

z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2..3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	09988	0.9989	0.9989	0.9989	0.9990	0.9990

Source: B. W. Lindgren, *Statistical Theory* (New York: Macmillan. 1962), pp. 392-393.