4. Probability and Probability Distributions

LPO.8800: Statistical Methods in Education Research

Sean P. Corcoran

Last time

Describing univariate distributions, cont.

- Measures of variability ("dispersion" or "spread")
- Measures of skewness
- Positional measures (e.g. percentiles, quartiles)
- Box plots, interquartile range

Last time

Data transformations:

- Linear transformations $(+, -, \times, \div)$
- z-scores
- Nonlinear transformations (e.g. square root, logarithm)
- Effects of transformations on the variable's distribution (mean, variability, shape)

Probability

In descriptive statistics, all of the outcomes are known—we are simply finding useful ways of summarizing them. When using statistics for inferential purposes, one has to think about *hypothetical* outcomes. Probability provides the tools to do this.

- Probability is a *model* of the **data generating process** (i.e., the process generating the outcomes we observe)
- This model tells us the relative frequency of outcomes in an unobserved *population*.
- Like any model, its usefulness depends on the application at hand, and its ability to provide a description of the real world.

Probability

Basic concepts:

- A **random process** is a realization of outcomes that are unknown in advance (*ex ante*) but where the possible outcomes are known.
- The **sample space** is the set of all possible outcomes of a random process. ("Everything that could happen").
- An **event** is a subset of the sample space. It may be a specific outcome or a collection (set) of outcomes.
- **Probability** is the likelihood of an event occurring. Often denoted $P(\text{event})$

Sample space

Examples of sample spaces:

- coin toss: {H, T}
- 2 coin tosses: {HH, HT, TH, TT}
- die roll: {1, 2, 3, 4, 5, 6}
- day of weather: {rain, no rain}
- amount of rain in a day $\{0 \rightarrow ?\}$
- years of education completed by an adult {1 - 20+}

The sample space may consist of *discrete* or *continuous* outcomes. It is sometimes denoted S.

Event

Examples of events:

- 2 coin tosses: {H on either toss}
- die roll: {2}, {1, 2}, or {even number}
- amount of rain in a day: {2 inches} or {1 to 3 inches}
- years of education completed by an adult: {16 or more}

We can use notation to refer to events. For example event $A = \{1, 2\}$ or $B = \{\text{Heads on either toss}\}$

Basic axioms of probability

There are three basic axioms, or postulates, of probability. They are the simplest of assumptions of any probability model:

1. $P(A) \geq 0$ for some event A
2. $P(S) = 1$ for the sample space S
3. If $A_1, A_2, ..., A_N$ are **mutually exclusive** events, then:

$$P(A_1 \cup A_2 \cup ... \cup A_N) = P(A_1) + P(A_2) + ... + P(A_N) = \sum_{i=1}^{N} P(A_i)$$

Probability rules: applying the axioms

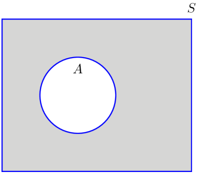
With **equally likely** mutually exclusive discrete outcomes, the probability of event A is:

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of possible outcomes}}$$

Examples: P(H) on a single coin toss $= 1/2$. P(2 *or* 3) on a die roll $= 2/6$.

Probability rules: applying the axioms

- If $\boldsymbol{P(A)}$ is the probability that event A occurs, $\boldsymbol{P(\sim A) = 1 - P(A)}$ is the probability that A does *not* occur.
- $\boldsymbol{\sim A}$ is the **complement** of A (something other than A occurs—grey area below)
- $\boldsymbol{P(A) + P(\sim A) = 1}$



Probability rules: applying the axioms

Examples:

- 2 coin tosses: $S = \{HH, HT, TH, TT\}$
  - P($\sim$coin comes up heads once) = $2/4 = 0.50$
  - P($\sim$two heads) = $(1 - (1/4)) = 3/4 = 0.75$

- die roll events: $\{2\}$, $\{1, 2\}$, or $\{$even number$\}$
  - P($\sim$2) = $(1 - (1/6)) = 5/6 = 0.833$
  - P($\sim$even number) = $3/6 = 0.50$

Probability rules: applying the axioms

- When A and B are two *mutually exclusive* events, then $\boldsymbol{P(A \cup B) = P(A) + P(B)}$

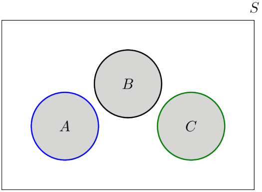
- $\cup$ is the symbol for the **union** of A and B—read as events A *or* B occurring.

- Mutually exclusive means there is no overlap in the events, so $\boldsymbol{P(A \cap B) = 0}$

- $\cap$ is the symbol for the **intersection** of A and B—read as events A and B *both* occurring.

Probability rules: applying the axioms

Mutually exclusive events A, B, and C:



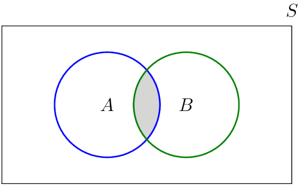
$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = 0$$

Probability rules: applying the axioms

- When A and B are *not* mutually exclusive events, then $\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

- That is, since A and B are overlapping, we must subtract the part common to both:



Non-mutually exclusive events: examples

Suppose you would like to enroll in an easy class (A) *or* a class that is graded easily (B). What is the probability a random class is *either* easy or graded easily?

- Assume P(A) = 0.25
- Assume P(B) = 0.35
- Assume $P(A \cap B) = 0.20$, that is 20% of classes are both easy *and* graded easily. Then:
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $\boldsymbol{P(A \cup B) = 0.25 + 0.35 - 0.20 = 0.40}$

Non-mutually exclusive events: examples

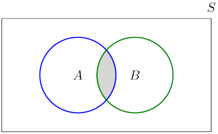
In the U.S. Census, race and ethnicity are separate questions. In 2010:

- 74.8% of the population is white (alone or in combination with others): $P(W) = 0.748$
- 16.3% of the population is Hispanic or Latino: $P(H) = 0.163$
- 9.5% of the population is white *and* Hispanic: $P(W \cap H) = 0.095$. Then:
- $P(W \cup H) = P(W) + P(H) - P(W \cap H)$
- $\boldsymbol{P(W \cup H) = 0.163 + 0.748 - 0.095 = 0.816}$

Conditional probability

Sometimes we are interested in knowing the probability than an event will occur *given that we know another event has occurred*. Suppose we have two events A and B. Associated with these events are 4 possibilities:

- $A \cap B$
- $A \cap \sim B$
- $\sim A \cap B$
- $\sim A \cap \sim$ B



Conditional probability

Now suppose we know B occurred. This means either of the following occurred:

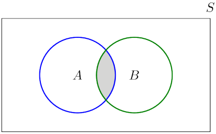
- $A \cap B$
- $\sim A \cap B$

In only one of these cases does A occur.

Conditional probability

Let $\boldsymbol{P(A|B)}$ be the **conditional probability** that event A occurs given that event B occurred.

$$P(A|B) = \frac{P(A \cap B)}{P(A \cap B) + P(\sim A \cap B)} = \frac{\boldsymbol{P(A \cap B)}}{\boldsymbol{P(B)}}$$



Probability rules: conditional probability

Example: In an election, 20 of 50 women and 12 of 40 men oppose a policy. A ballot drawn at random indicates opposition. What is the probability it was cast by a man? $P(\text{man}|\text{oppose})$

$$\begin{aligned} P(\text{man}|\text{oppose}) &= \frac{P(\text{man} \cap \text{oppose})}{P(\text{oppose})} \\ &= \frac{12/90}{32/90} \\ &= 0.375 \end{aligned}$$

What is the probability it was cast by a woman? $P(\text{woman}|\text{oppose})$

$$\begin{aligned} P(\text{woman}|\text{oppose}) &= \frac{P(\text{woman} \cap \text{oppose})}{P(\text{oppose})} \\ &= \frac{20/90}{32/90} \\ &= 0.625 \end{aligned}$$

Probability rules: conditional probability

Example: in a population of 100 people, 90 are vaccinated and 10 are not. 4 are infected with COVID-19, including 3 vaccinated and 1 unvaccinated. Among those infected, what is the likelihood they were vaccinated?

$$\begin{aligned} P(\text{vaccinated}|\text{infected}) &= \frac{P(\text{vaccinated} \cap \text{infected})}{P(\text{infected})} \\ &= \frac{3/100}{4/100} \\ &= 0.75 \end{aligned}$$

Seems alarming! But this conditional probability gets larger the greater the share of the population that is vaccinated.

Probability rules: conditional probability

Given a vaccination, what the probability of being infected with COVID-19? How does this compare to the unvaccinated?

$$\begin{aligned} P(\text{infected}|\text{vaccinated}) &= \frac{P(\text{vaccinated} \cap \text{infected})}{P(\text{vaccinated})} \\ &= \frac{3/100}{90/100} = 0.0333 \\ P(\text{infected}|\text{unvaccinated}) &= \frac{P(\text{unvaccinated} \cap \text{infected})}{P(\text{unvaccinated})} \\ &= \frac{1/100}{10/100} = 0.100 \end{aligned}$$

These conditional probabilities are much more informative about the benefits of vaccination.

Bayes' Rule

A useful application of conditional probabilities: using knowledge that B occurred to revise or "update" the probability that A occurred:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(B \cap A)}{P(B)} \\ &= \frac{\boldsymbol{P(B|A)P(A)}}{\boldsymbol{P(B)}} \end{aligned}$$

$P(A)$ is the *prior* and $P(A|B)$ is the *posterior*. The third line uses the definition of $P(B|A)$.

Bayes' Rule example

A common application of Bayes' Rule is in screening for a particular disease. For example, let the event A = person has breast cancer, and B = person has a positive mammogram. Assume:

- $P(A) = 1.4\%$
- $P(B) = 11.0\%$
- $P(B|A) = 78.6\%$ — mammogram is positive 78.6% of the time for someone with cancer

How worried should one be following a positive mammogram (in this example)? What is $P(A|B)$?

Bayes' Rule example

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.786 * 0.014}{0.11} = 0.100 = 10\%$$

In the population, 1.4% have breast cancer. Conditional on a positive mammogram, 10% have breast cancer. The new evidence (B) updates the probability of A.

Notably, the mammogram is positive 78.6% of the time for those with cancer ($P(B|A)$). Still, only 10% of those who screen positive have breast cancer ($P(A|B)$). Why? Breast cancer has a low *overall* incidence in the population ($P(A)$).

Independence

- If A and B are **independent events**, this means the probability that A occurs does not depend on whether B occurred, and vice versa.
- Put another way, knowing whether B occurred provides *no information* about the probability that event A occurred (and vice versa).
- When A and B are independent events, $P(A|B) = P(A)$
- From the definition of conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
- Thus when A and B are independent, $\boldsymbol{P(A \cap B) = P(A) \times P(B)}$. This is the *multiplicative rule*.

Independence: examples

Are these independent events?

- Rolling a die: event A is "rolling an even number," and event B is "rolling a 2 or 4." Is it true that $P(A \cap B) = P(A) \times P(B)$?
- Tossing a coin twice: event A is "getting H on the first toss," and event B is "getting H on the second toss."
- Drawing a random person from the U.S. Census: event A is "the person is white," and event B is "the person is Hispanic."
- Event A is having breast cancer; event B is having a positive mammogram.
- Event A is having had COVID-19 vaccination; event B is having a COVID infection.

Independence: examples

The outcomes of sampling with replacement are independent events. Recall that 16.3% of the population in the 2000 Census was Hispanic.

- Draw three random people from the Census. What is the probability that all three are Hispanic?
- $\boldsymbol{P(HHH) = 0.163 \times 0.163 \times 0.163} = 0.0043$
- Draw three random people from the Census. What is the probability that the first is Hispanic but the other two are not?
- $\boldsymbol{P(H, \sim H, \sim H) = 0.163 \times (1 - 0.163) \times (1 - 0.163) = 0.114}$

Conceptions of probability

Where do probabilities come from?

- **Classical interpretation**: the assumption of equally likely outcomes provides a logical assignment of probability. Ex: P(H) on a single coin toss is 0.50. P(2 *or* 3) on a die roll will be $2/6 = 1/3$.

- **Frequentist interpretation**: if a random trial is repeated a large (infinite) number of times under identical conditions, $P(A)$ is the fraction of times A occurs.
  - We might assume the probability that a newborn baby is female is 0.5 (or 50%). In fact, over millions of repeated observations, it is closer to 0.48 in the U.S.

- **Subjective interpretation**: probabilities based on intuition

Probability as relative frequency

Probabilities as relative frequencies in the population: consider the Maryland Pick 3 lottery

- \$1 to play

- Choose three numbers between 0 and 9 (e.g. 000 $\rightarrow$ 999)

- Win \$500 if the three numbers are picked exactly

Probability as relative frequency

Consider the probability of certain events:

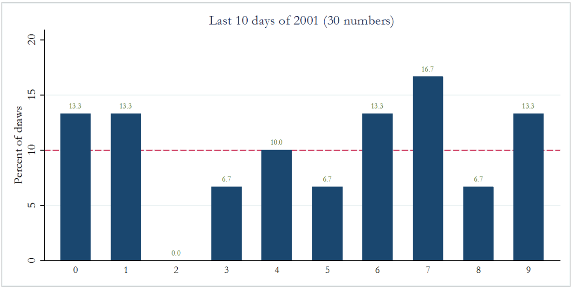
- Theoretical probabilities: how often would one *expect* to see a particular integer come up in one of the positions? (1/10)

- Probability as relative frequency: how often does a number *actually come up?*

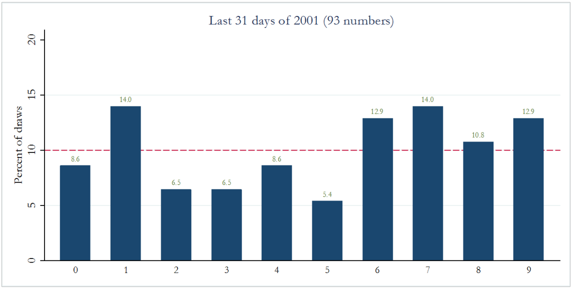
Probability as relative frequency

As the number of draws increases, the observed relative frequency should approach the true probabilities (assuming the game is fair).

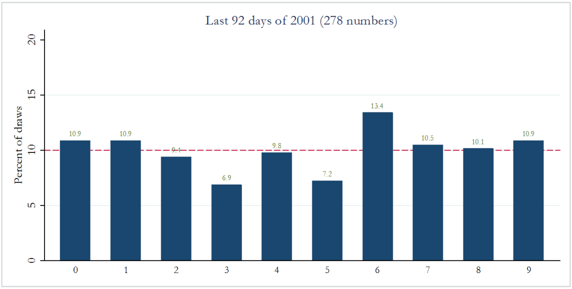
Probability as relative frequency



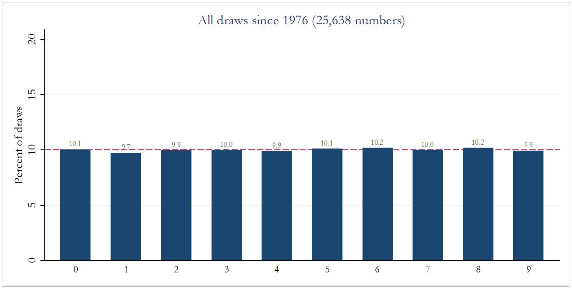
Probability as relative frequency



Probability as relative frequency



Probability as relative frequency



Probability distributions

A **probability distribution** (or probability density) allows one to describe all of the possible outcomes of a random process and the probabilities those outcomes occur.

The form of the probability distribution function (PDF) depends on whether the outcome is *discrete* or *continuous*.

Discrete probability distributions

For a **discrete** outcome, the probability distribution is a list of all possible outcomes and their probabilities. These probabilities must be positive and sum to 1.

- Analogous to the relative frequency distribution from descriptive stats.
- Simple example: # of heads in 3 coin flips:

X	P(X)
0	1/8
1	3/8
2	3/8
3	1/8
SUM	1

Discrete probability distributions

The probability that X is between a and b, inclusive, is:

$$P(a \leq X \leq b) = \sum_{i=a}^{b} P(X = i)$$

For example, using the PDF above:

$$P(2 \leq X \leq 3) = P(X = 2) + P(X = 3) = (3/8) + (3/8) = 0.75$$

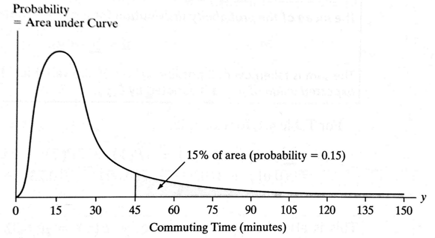
Continuous probability distributions

For **continuous** outcomes, a probability distribution function assigns probabilities to *intervals* of values.

- Continuous random variables take on *infinitely many* outcomes
- The probability such a variable takes on a *specific* value is **zero**.
- The probability of the outcome falling in the interval containing *all* possible values is **one**.

Continuous probability distributions

When graphed, the **area** under a continuous probability distribution represents the probability that an outcome will fall within a particular interval:



Continuous probability distributions

Some will recognize that the area under a continuous function $f(x)$ between a and b can be found as:

$$\int_a^b f(x)dx$$

The area under $f(x)$ from $-\infty$ to $+\infty$ must be 1 to be a valid PDF.

Depending on the PDF, these areas can be obtained from formulas, tables, calculators, Stata, etc.

What to know about probability distribution functions

- How the function assigns a probability to every possible value of x
- "First principles": what kind of a process lends itself to such a distribution? How does the distribution connect to the real world and the substantive problem at hand?
- How to use the final expression for the PDF.
- How to *simulate* random draws from the PDF.
- How to calculate features of the distribution such as its expected value/mean and variance (its **moments**).

Discrete probability distribution: Bernoulli

A **Bernoulli** distribution is the simplest discrete PDF, for a 0-1 outcome like the gender of a newborn child. If $P(X) = P(boy) = 0.514$:

X	$P(X)$
0	0.486
1	0.514

Generally if $P(X = 1) = \pi$, then: $P(X) = \pi^X (1 - \pi)^{(1-X)}$ for $X = 0, 1$

Discrete probability distribution: binomial

Consider the possible gender combinations of 3 children had in sequence, where the order matters and births are independent events. P(B)=0.514 and P(G)=0.486.

Sequence	Probability
BBB	0.136
BBG	0.128
BGB	0.128
GBB	0.128
BGG	0.121
GGB	0.121
GBG	0.121
GGG	0.115
SUM	1.000

Discrete probability distributions: binomial

The above PDF can be written another way to show the probability of having k boys in 3 births:

# of boys	Probability
0	0.1148
1	0.3642
2	0.3852
3	0.1358
SUM	1.000

This is an example of a **binomial distribution**.

Discrete probability distributions: binomial

The binomial distribution is applicable when one has repeated *independent* and *identical* trials where the outcome of each trial is Bernoulli.

Define $X = 1$ generically as "success". If $\pi = P(X = 1)$ on each trial, then the probability of k successes in n trials is:

$$P(k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

where:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the number of ways to select k elements from a possible of n objects where the order does not matter. Note $n! = n * (n-1) * (n-2) * ... * 1$

Expected value: discrete random variable

The **mean** or **expected value** of a discrete random variable X with N unique outcomes is denoted $E(X)$ or μ_X and given by:

$$E(X) = \sum_{i=1}^{N} x_i P(x_i)$$

That is, it is the weighted average of outcomes where the weights are the probabilities each outcome occurs. It is a measure of *central tendency*.

In the example above of 3 births:

$$E(X) = (0 \times 0.1148) + (1 \times 0.3642) + (2 \times 0.3852) + (3 \times 0.1358) = 1.542$$

The mean number of boys in 3 births is 1.542.

Variance: discrete random variable

The **variance** of a discrete random variable X with N unique outcomes is denoted $Var(X)$ or σ_X^2 and given by:

$$Var(X) = E(X - E(X))^2 = \sum_{i=1}^{N}(x_i - E(X))^2 P(x_i)$$

That is, it is the weighted average of squared deviations of each outcome X_i from $E(X)$, where the weights are the probabilities each outcome occurs. It is a measure of *dispersion*.

Mean and variance: Bernoulli

Recall the Bernoulli PDF $P(X) = \pi^X(1-\pi)^{(1-X)}$ where X is either 0 or 1.

$$E(X) = 0 * P(0) + 1 * P(1) = \boldsymbol{\pi}$$

In our newborn gender case, $E(X) = 0.514$. The mean value of X is 0.514. (Or, 51.4% of newborns in the population are boys).

$$Var(X) = E(X - E(X))^2 = \Big[(0-\pi)^2 * (1-\pi)\Big] + \Big[(1-\pi)^2 * (\pi)\Big] = \boldsymbol{\pi(1-\pi)}$$

Mean and variance: binomial

Recall the binomial PDF $P(k) = \binom{n}{k}\pi^k(1-\pi)^{n-k}$.

$$E(X) = \boldsymbol{n\pi}$$

On average, in n trials there are $n\pi$ successes. In our 3 successive births example, on average a mother will have $3 * 0.514 = 1.542$ boys. (Same answer as earlier).

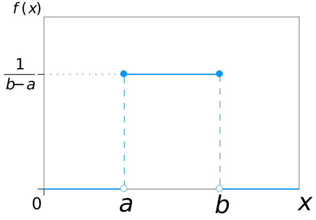
$$Var(X) = \boldsymbol{n\pi(1-\pi)}$$

The variance in the number of boys (out of 3 successive births) will be $3 * 0.514 * (1 - 0.514) = 0.749$

The uniform distribution

The **uniform (or rectangular) distribution** is a common continuous probability distribution function. Here:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$



The uniform distribution

This is a valid PDF since the area under this function between a and b is 1. As a continuous PDF, the probability that X falls between values c and d (assuming $a \leq c \leq d \leq b$) is:

$$\int_c^d \frac{1}{b-a} dx$$

For the uniform distribution this is just the area of a rectangle:

$$(d - c) * \frac{1}{b-a}$$

Expected value: continuous random variable

The **mean** or **expected value** of a continuous random variable X with PDF $f(X)$ is denoted $E(X)$ or μ_X and given by:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

The interpretation is analogous to the mean of a discrete random variable: it is a “weighting” of the possible x’s, using the probability density as weights.

Variance: continuous random variable

The **variance** of a continuous random variable X with PDF $f(X)$ is denoted $Var(X)$ or σ_X^2 and given by:

$$Var(X) = E(X - E(X))^2 = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx$$

Mean and variance: uniform

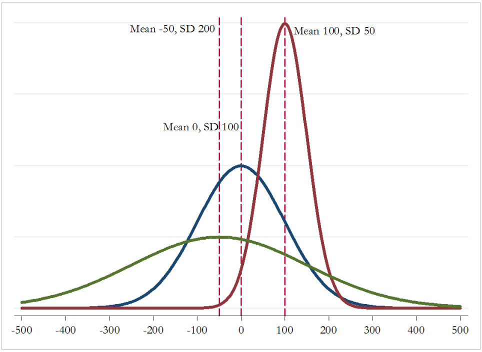
$$E(X) = \frac{1}{2}(a + b)$$

$$Var(X) = \frac{1}{12}(b - a)^2$$

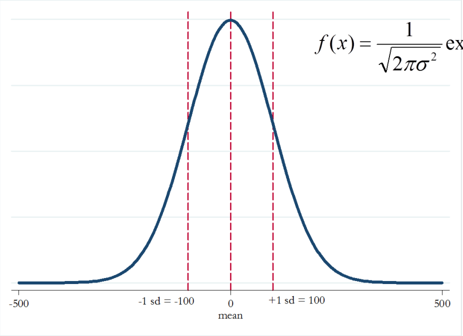
The normal distribution

- The **normal distribution** is probably the most important probability distribution in statistics (aka "Gaussian")
- It is really a *family* of distributions, with different population means (μ) and standard deviations (σ)
- The normal distribution is *symmetric*: $P(x > \mu) = P(x < \mu) = 0.50$
- The area under the normal distribution is **1**.

The normal distribution



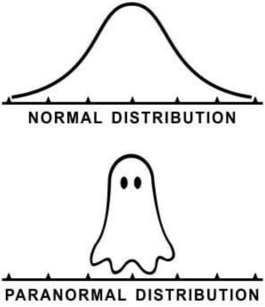
The normal distribution



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right)$$

The normal distribution

The paranormal distribution



The normal distribution

When a variable X has a normal distribution:

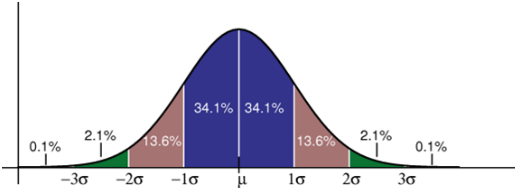
- The area between two values a and b is the probability that a randomly drawn X will fall within that interval: $P(a < X < b)$

- The area from $-\infty$ to k is the probability that a randomly drawn X will fall *below* k: $P(X < k)$

- The area from k to $+\infty$ is the probability that a randomly drawn X will fall *above* k: $P(X > k)$

The normal distribution

Recall the empirical rule: holds for all normal distributions



The normal distribution

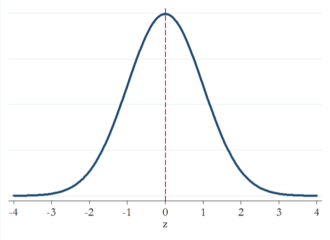
- Probabilities from the normal distribution can be obtained using Stata or an online calculator. Historically, they were taken from tables (as in the back of your text). You should know how to obtain probabilities using multiple methods.

- How would a table of probabilities work? As we saw, there are as many normal distributions as there are values of μ and σ.

- Solution: all normally distributed variables can be converted to **standard normal**—a *z*-score

The standard normal distribution

If X has a normal distribution, then its z-score: $z = (X - \mu)/\sigma$ has a **standard normal distribution** with a mean of zero and s.d. of 1:



The standard normal distribution

The standard normal makes computing probabilities easier:

$$P(X > x_1) = P\left(\frac{X - \mu}{\sigma} > \frac{x_1 - \mu}{\sigma}\right) = P\left(z > \frac{x_1 - \mu}{\sigma}\right)$$

The standard normal distribution: example

Suppose X has a normal distribution with a mean of 5 and a standard deviation of 2. What is the probability $X < 7.5$?

- It is common to write: $X \sim N(\mu, \sigma)$ or here $X \sim N(5, 2)$
- $P(X < 7.5) = P((X - 5)/2 < (7.5 - 5)/2)$
- Note $(7.5 - 5)/2 = 1.25$ is the z-score corresponding to $X = 7.5$ in the original distribution. 7.5 is 1.25 standard deviations above the mean.
- $\boldsymbol{P(z < 1.25) = 0.8944}$ using an online calculator, Stata, or a lookup table.

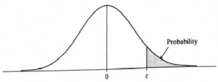
The standard normal distribution: example

To compute probabilities from the standard normal distribution in Stata:

- The function `normal(z)` provides the probability of falling below z (the *cumulative probability*, or left tail)
- Example: `display normal(1.96)` yields 0.9750021
- Note `display 1-normal(z)` provides the probability of falling *above* z, the right tail

The standard normal distribution

A typical standard normal lookup table:

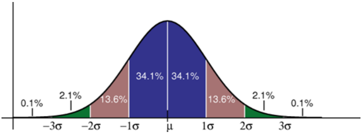


	Second Decimal Place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	[illegible]	[illegible]	[illegible]	[illegible]					

The standard normal distribution

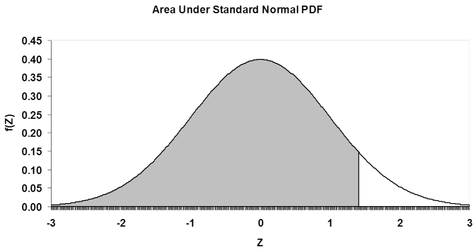
Verifying the Empirical Rule using a standard normal distribution: $X \sim N(0,1)$

- $P(z > 2) = P(z < -2) = 0.021 + 0.001 = 0.022$
- $P(-1 < z < 1) = 0.341 + 0.341 = 0.682$
- $P(-2 < z < 2) = 0.136 + 0.341 + 0.341 + 0.136 = 0.954$



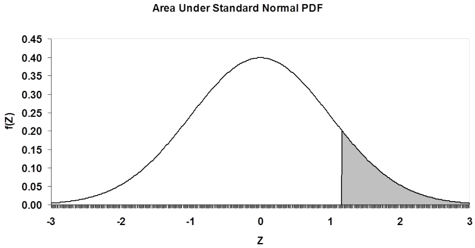
The standard normal distribution: examples

$P(z < 1.41) = 0.9207$



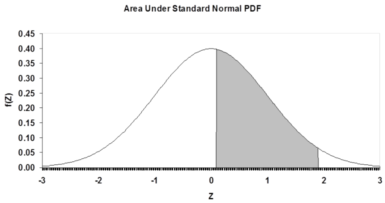
The standard normal distribution: examples

$P(z > 1.17) = 1 - P(z < 1.17) = 1 - 0.8790 = 0.1210$



The standard normal distribution: examples

$P(0.1 < z < 1.9) = P(z < 1.9) - P(z < 0.1) = 0.9713 - 0.5398 = 0.4315$



The standard normal distribution: examples

The time to complete an exam (X) is normally distributed with a mean of 110 and a standard deviation of 20. What is the fraction of students who will *not* finish in 120 minutes?

- $X \sim N(110, 20)$
- $P(X > 120) = 1 - P(X < 120)$
- $P(X < 120) = P((X - 110)/20 < (120 - 110)/20)$
- $1 - P(z < 0.5) = 1 - 0.6915 = 0.3085$

The standard normal distribution

It is also possible to find the z value corresponding to a particular tail probability:

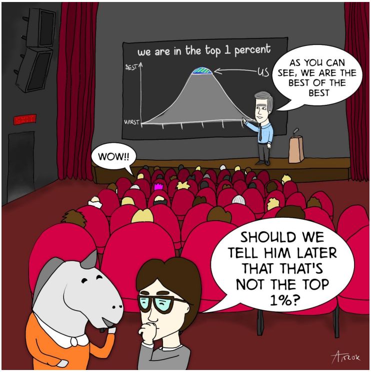
- The function `invnormal(p)` provides the z value for which the probability of falling *below* z is p
- Example: `display invnormal(0.025)` yields -1.96
- The probability of falling 1.96σ or more below the mean *or* 1.96σ or more above the mean is $2 \times 0.025 = \mathbf{0.05}$

The standard normal distribution: examples

The time to complete an exam (X) is normally distributed with a mean of 110 and a standard deviation of 20. Above what value will X fall 25% of the time?

- $X \sim N(110, 20)$
- First find the value above which z will fall 25% of the time ($z = 0.6745$)
- The value of X that is 0.6745 standard deviations above its mean is: $110 + (0.6745 \times 20) = 123.49$

The normal distribution



Simulating draws from probability distributions in Stata

Random number generators—like the one in Stata—are really *pseudo*-random number generators. That is, they are perfectly predictable mathematical functions.

If you know the pseudo-random number generator and *seed*, you will get the same sequence of numbers every time.

In Stata, `set seed #` where # is a number you choose.

This will allow you (and others) to replicate your random draws.

Simulating draws from probability distributions in Stata

Start from an empty dataset, and choose the number of observations you wish to generate:

```
clear
set seed 1989
set obs 1000
```

To generate 1000 draws from the uniform(0,1) distribution:

```
gen x1=runiform()
```

To generate 1000 draws from the uniform(a,b) distribution:

```
gen x2=runiform(a,b)
```

To generate 1000 draws from the binomial(n,p) distribution—returns integers:

```
gen x3=rbinomial(n,p)
```


Simulating draws from probability distributions in Stata

To generate 1000 draws from the normal(μ, σ) distribution:

```
gen x4=rnormal(m, s)
```

To generate 1000 draws from the standard normal distribution:

```
gen x5=rnormal()
```

To generate 1000 draws from the Bernoulli distribution, use the uniform(0,1) to generate random numbers and then assign the Bernoulli value (0 or 1) based on the result. For example, if $\pi = 0.514$:

```
gen x6=runiform()
gen x7=1 if x6<0.514
replace x7=0 if x6>=0.514
```