Chapter 4: Random Variables and Distributions

(Ott & Longnecker Sections: 4.6-4.10)

Duzhe Wang

https://dzwang91.github.io/stat371/

Part 2

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

What do we study?

We'll continue RV.

Key Concepts: See outline.

Outline

An example: toss of coin

If we toss a fair coin, we define X be the number of head up, then $X = 1$ if head comes up and $X = 0$ if tail comes up. Because this is a fair coin, then the pmf would be:

An example: toss of coin

If we toss a fair coin, we define X be the number of head up, then $X = 1$ if head comes up and $X = 0$ if tail comes up. Because this is a fair coin, then the pmf would be:

x	$p(x)$
0	1/2
1	1/2

An example: toss of coin

If we toss a fair coin, we define X be the number of head up, then $X = 1$ if head comes up and $X = 0$ if tail comes up. Because this is a fair coin, then the pmf would be:

x	$p(x)$
0	$1/2$
1	$1/2$

This RV is discrete because it can only take values 0 and 1. Note that the total of the probabilities is 1, and this will be the case for every RV.

Example continued

Let's discuss a bit of notation.

Example continued

Let's discuss a bit of notation.

What does $p(0) = \frac{1}{2}$ mean?

Example continued

Let's discuss a bit of notation.

What does $p(0) = \frac{1}{2}$ mean?

- $P(X = 0) = 1/2$

Example continued

Let's discuss a bit of notation.

What does $p(0) = \frac{1}{2}$ mean?

- $P(X = 0) = 1/2$
- We should read this in words as, "the probability that the RV X realizes to the value 0 is $1/2$."

Bernoulli RV

The RV X in the example is a special RV:

- We call a RV a Bernoulli RV if it can only realize to the values 0 or 1. The probability that it realizes to 1 is called π. A Bernoulli RV X can be denoted as $X \sim Bern(\pi)$. The symbol "$\sim$" should be read "distributed as."

In the example toss of coin, $\pi = 1/2$, so we can say, $X \sim Bern(1/2)$.

Outline

A natural generalization

Now if we toss a fair coin 10 times, let X be the number of heads, how do we characterize this random variable X?

Binomial random process

A binomial random process has the following properties:

- The random process consists of n identical Bernoulli variables which we call trials.
- In each Bernoulli trial we call an outcome of 1 a success, and an outcome of 0 a failure.
- The probability of a success on any single trial is the same for every trial, and is denoted π.
- The trials are independent, in that the outcome of any trial does not affect the outcome of any other trial.

Binomial random process examples

For each of the following questions, say whether the random process is a binomial process or not.

Binomial random process examples

For each of the following questions, say whether the random process is a binomial process or not.

- One basketball player attempts 10 free throws and the number of successful attempts is recorded.

Binomial random process examples

For each of the following questions, say whether the random process is a binomial process or not.

- One basketball player attempts 10 free throws and the number of successful attempts is recorded.

The process does consist of trials(each attempt), and the result has two outcomes, success and failure. However, it is unlikely that the trials are independent. The performance on the previous attempt might affect the probability of making the next attempt. The player might, for example, concentrate harder after a miss. Or they might get discouraged after a miss and concentrate less. Therefore this is not a good approximation to a binomial process.

Binomial random process examples

- Ten different basketball players each attempt 1 free throw and the total number of success attempts is recorded.

Binomial random process examples

- Ten different basketball players each attempt 1 free throw and the total number of success attempts is recorded.

The process does consist of trials(each attempt), the result has two outcomes, success and failure and as long as the players do not influence one another, they are probably approximate independent. However, it is unlikely that every player has the same chance of making a free throw, therefore the probability of success on each trial is not constant. So this is not a good approximation to a binomial process.

Outline

Binomial RV

If the random process is a binomial random process, we then define:

- A binomial RV, call it B, is the total number of successes achieved in n trials of a binomial random process with probability π of success on any given trial. We denote such an RV as $B \sim Bin(n, \pi)$.

Binomial RV

If the random process is a binomial random process, we then define:

- A binomial RV, call it B, is the total number of successes achieved in n trials of a binomial random process with probability π of success on any given trial. We denote such an RV as $B \sim Bin(n, \pi)$.

A Binomial RV can be thought of as the sum of n independent Bernoulli RVs that all have the same probability of success π. Thus, the values a $Bin(n, \pi)$ can take are $0, 1, \ldots, n$.

Example: manufacture of circuit boards

Suppose, based on several years of testing, it is determined that 96% of circuit boards are fully operational. A warehouse contains a very large population of boards. If we select 5 boards at random and then record the number of operational boards in that sample of 5,

- Can this process be described by a binomial process?

Example: manufacture of circuit boards

Suppose, based on several years of testing, it is determined that 96% of circuit boards are fully operational. A warehouse contains a very large population of boards. If we select 5 boards at random and then record the number of operational boards in that sample of 5,

- Can this process be described by a binomial process?
- Let $Y =$ the number of operational boards in that sample of 5, then

Example: manufacture of circuit boards

Suppose, based on several years of testing, it is determined that 96% of circuit boards are fully operational. A warehouse contains a very large population of boards. If we select 5 boards at random and then record the number of operational boards in that sample of 5,

- Can this process be described by a binomial process?
- Let $Y =$ the number of operational boards in that sample of 5, then

$$Y \sim Bin(5, 0.96)$$

Pmf of Binomial RV

For a binomial RV $B \sim Bin(n, \pi)$, the probability of observing b successes is:

$$p(b) = p(B = b) = \frac{n!}{b!(n-b)!}\pi^b(1-\pi)^{n-b},$$

where $n!$ is called the factorial, and is the product of all integers from n to 1. By definition, $0! = 1$.

Pmf of Binomial RV

For a binomial RV $B \sim Bin(n, \pi)$, the probability of observing b successes is:

$$p(b) = p(B = b) = \tfrac{n!}{b!(n-b)!}\pi^b(1-\pi)^{n-b},$$

where $n!$ is called the factorial, and is the product of all integers from n to 1. By definition, $0! = 1$.
For the circuit board example, we could calculate the probability that the RV Y realizes to 4 (i.e. the chance 4 of the 5 boards are working) as:

$$p(4) = P(Y = 4) = \tfrac{5!}{4!(5-4)!}0.96^4(1-0.96)^{5-4} = 0.17.$$

Outline

Expectation of a RV

The following are important properties of random variables we will see many times in this course.

The expectation or expected value of a RV X, denoted $E(X)$ or μ_X, is the mean of the population. The expectation of a *discrete* RV X is:

$$\mu_X = E(X) = \sum_x x * p(x)$$

where the sum is taken over all possible realizations of X. (Note that we can define the expected value for continuous RVs using the pdf and *integrals*, which allow you to take continuous sums.)

Variance and standard deviation of a RV

- The variance of a RV X, denoted $VAR(X)$, or σ_X^2 is the variance of the population. The variance of a *discrete* RV X is:

$$\sigma_X^2 = VAR(X) = \sum_x p(x) * (x - E(X))^2.$$

Variance and standard deviation of a RV

- The variance of a RV X, denoted $VAR(X)$, or σ_X^2 is the variance of the population. The variance of a *discrete* RV X is:

$$\sigma_X^2 = VAR(X) = \sum_x p(x) * (x - E(X))^2.$$

- Standard deviation of a RV: $SD(X) = \sqrt{VAR(X)}$. It happens that sometimes it's easier to work with standard deviations, and sometimes it's easier to work with variances.

Example

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

Example

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

- $E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$

Example

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

- $E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$
- $VAR(X) = (1/2) * (0 - 1/2)^2 + (1/2) * (1 - 1/2)^2 = 1/4$

Example

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

- $E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$
- $VAR(X) = (1/2) * (0 - 1/2)^2 + (1/2) * (1 - 1/2)^2 = 1/4$
- $SD(X) = \sqrt{VAR(x)} = 1/2$

Example

Let's compute the expectation, variance and standard deviation of the RV in the example toss of coin, $X \sim Bern(1/2)$:

- $E(X) = 0 * (1/2) + 1 * (1/2) = 1/2$
- $VAR(X) = (1/2) * (0 - 1/2)^2 + (1/2) * (1 - 1/2)^2 = 1/4$
- $SD(X) = \sqrt{VAR(x)} = 1/2$

What do you find?

Properties

For a general Bernoulli RV $X \sim Bern(\pi)$, we have $E(X) = \pi$ and $VAR(X) = \pi(1 - \pi)$.

Properties

For a general Bernoulli RV $X \sim Bern(\pi)$, we have $E(X) = \pi$ and $VAR(X) = \pi(1 - \pi)$.

What about for binomial RV?

Properties

For a general Bernoulli RV $X \sim Bern(\pi)$, we have $E(X) = \pi$ and $VAR(X) = \pi(1-\pi)$.

What about for binomial RV?

It can be shown that

if $B \sim Bin(n, \pi)$, then $E(B) = n\pi$, and $VAR(B) = n\pi(1-\pi)$.

Example

Back to the circuit boards. Recall that Y denotes the number of operational boards in five random boards, and $Y \sim Bin(5, 0.96)$. Then

$$E(Y) = 5 * 0.96 = 4.8$$
$$Var(Y) = 5 * 0.96 * 0.04 = 0.192$$

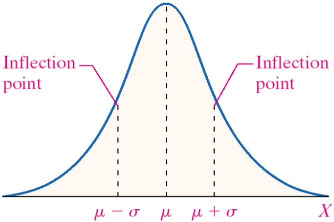
Outline

Continuous RV

Recall that continuous RVs take values in specified ranges, and their probability distributions are called probability density functions, or pdfs, which are denoted $f(x)$. The area under the curve described by the pdf between any two possible realizations of the RV determines the probability that the RV will realize to a value in that range.

Outline

The bell curve



It's a special curve, called a normal, Gaussian, or bell curve. If the pdf of X is a bell curve, then we say X has the normal distribution. Lots of biological random variables are normal, for example, body weight, crop yield, protein content in soybean, density of blood components.

Outline

Normal distribution

If X is a normal RV, then the pdf is

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

Normal distribution

If X is a normal RV, then the pdf is

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

- $\mu = E(X), \sigma^2 = VAR(X)$. We often denote a normal RV X with mean μ and variance σ^2 by $X \sim N(\mu, \sigma^2)$.

Normal distribution

If X is a normal RV, then the pdf is

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

- $\mu = E(X), \sigma^2 = VAR(X)$. We often denote a normal RV X with mean μ and variance σ^2 by $X \sim N(\mu, \sigma^2)$.
- Don't confuse this π, which is the fundamental math constant that is approximately 3.14, with the π that represents the chance of success in a Bernoulli trial!

Normal distribution

If X is a normal RV, then the pdf is

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

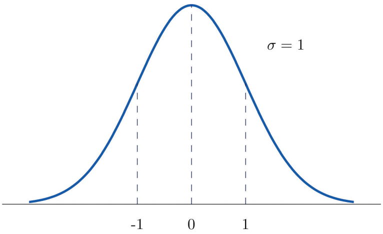
- $\mu = E(X), \sigma^2 = VAR(X)$. We often denote a normal RV X with mean μ and variance σ^2 by $X \sim N(\mu, \sigma^2)$.
- Don't confuse this π, which is the fundamental math constant that is approximately 3.14, with the π that represents the chance of success in a Bernoulli trial!
- You can play with the bell curve at https://dzwang.shinyapps.io/thebellcurve/

Properties of normal distribution

1. A normal RV can theoretically realize to any value between $-\infty$ and ∞.
2. The normal distribution is symmetric around the mean, μ.
3. The inflection points (points where the curve moves from concave downward to concave upward) are at $\mu \pm \sigma$.
4. The total area under the curve is 1.
5. The area under the curve between $\mu - \sigma$ and $\mu + \sigma$ is about 0.68; the area under the curve between $\mu - 2\sigma$ and $\mu + 2\sigma$ is about 0.95. Very little of the area is farther than three sds from the mean (about 0.003).

Outline

The standard normal distribution



We call the $N(0, 1)$ distribution the standard normal distribution and we usually reserve Z to denote a standard normal RV.

The standard normal distribution

- $P(Z \leq 0) =?$
- $P(Z = 0) =?$
- $P(Z < 1) =?$
- $P(0 \leq Z \leq 1) =?$
- $P(-1 \leq Z \leq 1) =?$
- $P(Z > 1.5) =?$

The standard normal distribution

- $P(Z \leq 0) =?$
- $P(Z = 0) =?$
- $P(Z < 1) =?$
- $P(0 \leq Z \leq 1) =?$
- $P(-1 \leq Z \leq 1) =?$
- $P(Z > 1.5) =?$

$$P(Z \leq 0) = \frac{1}{2} \text{ (by symmetry)}$$

The standard normal distribution

- $P(Z \leq 0) = ?$
- $P(Z = 0) = ?$
- $P(Z < 1) = ?$
- $P(0 \leq Z \leq 1) = ?$
- $P(-1 \leq Z \leq 1) = ?$
- $P(Z > 1.5) = ?$

$$P(Z \leq 0) = \tfrac{1}{2} \text{ (by symmetry)}$$

$$P(Z = 0) = 0 \text{ (because Z is continuous)}$$

The standard normal distribution

- $P(Z \leq 0) = ?$
- $P(Z = 0) = ?$
- $P(Z < 1) = ?$
- $P(0 \leq Z \leq 1) = ?$
- $P(-1 \leq Z \leq 1) = ?$
- $P(Z > 1.5) = ?$

$$P(Z \leq 0) = \tfrac{1}{2} \text{ (by symmetry)}$$

$$P(Z = 0) = 0 \text{ (because Z is continuous)}$$

But how do we calculate others?

Outline

Z table

Standard Normal Probabilities

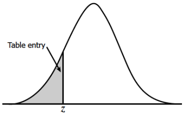


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148

Z table

Standard Normal Probabilities

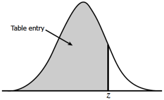


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

$$P(Z < 1) = ?$$

Z table

Standard Normal Probabilities

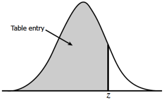


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

$$P(Z < 1) = ?$$

0.8413

Z table

$$P(0 \leq Z \leq 1) = ?$$

Z table

$$P(0 \leq Z \leq 1) = ?$$

$$P(0 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq 0) = 0.8413 - 0.5 = 0.3413$$

Z table

$$P(0 \leq Z \leq 1) = ?$$

$$P(0 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq 0) = 0.8413 - 0.5 = 0.3413$$

$$P(-1 \leq Z \leq 1) = ?$$

Z table

$$P(0 \leq Z \leq 1) =?$$

$$P(0 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq 0) = 0.8413 - 0.5 = 0.3413$$

$$P(-1 \leq Z \leq 1) =?$$

$$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) = 0.8413 - 0.1587 = 0.6826$$

Z table

$$P(0 \leq Z \leq 1) = ?$$

$$P(0 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq 0) = 0.8413 - 0.5 = 0.3413$$

$$P(-1 \leq Z \leq 1) = ?$$

$$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) = 0.8413 - 0.1587 = 0.6826$$

$$P(Z > 1.5) = ?$$

Z table

$$P(0 \leq Z \leq 1) = ?$$

$$P(0 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq 0) = 0.8413 - 0.5 = 0.3413$$

$$P(-1 \leq Z \leq 1) = ?$$

$$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) = 0.8413 - 0.1587 = 0.6826$$

$$P(Z > 1.5) = ?$$

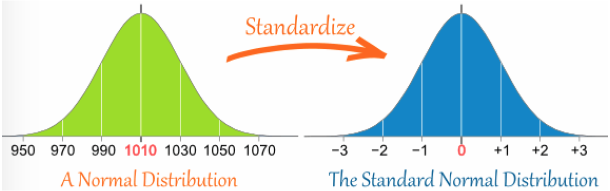
$$P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.9332 = 0.0668$$

But what if...

$Y \sim N(3, 0.25^2)$, how do we calculate $P(Y \leq 2.8)$?

But what if...

$Y \sim N(3, 0.25^2)$, how do we calculate $P(Y \leq 2.8)$?



Outline

Standardization

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Note: this is a very important technique, we'll use this fact a lot of times in this course.

Standardization

Example. $Y \sim N(3, 0.25^2)$, how do we calculate $P(Y \leq 2.8)$?

Standardization

Example. $Y \sim N(3, 0.25^2)$, how do we calculate $P(Y \leq 2.8)$?

- Step 1:

$$P(Y \leq 2.8) = P(\tfrac{Y-3}{0.25} \leq \tfrac{2.8-3}{0.25}),$$

then we have

$$P(Y \leq 2.8) = P(Z \leq -0.8).$$

Example. $Y \sim N(3, 0.25^2)$, how do we calculate $P(Y \leq 2.8)$?

- Step 1:

$$P(Y \leq 2.8) = P(\tfrac{Y-3}{0.25} \leq \tfrac{2.8-3}{0.25}),$$

  then we have

$$P(Y \leq 2.8) = P(Z \leq -0.8).$$

- Step 2: Use the Z table to compute $P(Z \leq -0.8)$, and we get 0.21.

But what if...

Example. $Y \sim N(3, 0.25^2)$, how do we find y such that

$$P(Y \geq y) = 0.25$$

Outline

Reverse standardization

If $Z \sim N(0,1)$, then $X = Z\sigma + \mu \sim N(\mu, \sigma^2)$.

Note: this is also a very important technique.

z critical value

Let $Z \sim N(0, 1)$, α is given, then the value z such that $P(Z \geq z) = \alpha$ is called z_α. We call z_α a *z critical value*. It can be thought of as the $1 - \alpha$ quantile of the standard normal distribution (recall the definition of quantile from the section on descriptive statistics!).

Example

$Y \sim N(3, 0.25^2)$, how do we find y such that

$$P(Y \geq y) = 0.25$$

Example

$Y \sim N(3, 0.25^2)$, how do we find y such that

$$P(Y \geq y) = 0.25$$

- Step 1:

$$\text{Find } z \text{ such that } P(Z \geq z) = 0.25,$$

  Use Z table, we find the 0.25 critical value of the standard normal distribution to be $z_{0.25} = 0.67$.

Example

$Y \sim N(3, 0.25^2)$, how do we find y such that

$$P(Y \geq y) = 0.25$$

- Step 1:

  $$\text{Find } z \text{ such that } P(Z \geq z) = 0.25,$$

  Use Z table, we find the 0.25 critical value of the standard normal distribution to be $z_{0.25} = 0.67$.

- Step 2:

  Use reverse standardization to find the appropriate value of y. Reverse standardizing gives $0.25 * 0.67 + 3 = 3.17$.

A reminder

You need to know how to use Z table to calculate probability and find the z critical value, particularly for exams. It's very IMPORTANT, so feel free to do many different examples until you feel confident and comfortable.

Doing calculations with R

```
> pnorm(1)
[1] 0.8413447
> pnorm(2)-pnorm(-2)
[1] 0.9544997
> pnorm(3)-pnorm(-3)
[1] 0.9973002
> 1- pnorm(137, mean=112, sd=10)
[1] 0.006209665
> qnorm(.95, mean=112, sd=10)
[1] 128.4485
> pbinom(1, size=6, prob=1/6)
[1] 0.7367755
> dbinom(0:6, size=6, prob=1/6)
[1] 0.335 0.402 0.201 0.054 0.008 0.001 0.000
```

`norm`	normal distribution
`binom`	binomial
`p`	probability: $\mathbb{P}\{Y \leq \dots\}$
`q`	quantile
`d`	density, or probability mass function: $\mathbb{P}\{Y = \dots\}$

What's the next?

In the next lecture, we'll discuss the distributions of functions of RVs and concepts of estimation.