STA257: Probability and Statistics 1

Instructor: Katherine Daignault

Department of Statistical Sciences
University of Toronto

Week 5

Outline

Outline

Calculus Refresher - Derivatives

- This week isn't too calculus heavy - you will just need to remember derivatives.
- Notation: the first derivative of a function $f(x)$ is denoted by $f'(x) = \frac{d}{dx}f(x)$
- Some useful results:
  - if $f(x) = x^r$ then $f'(x) = rx^{r-1}$
  - $f(x) = e^x = f'(x)$ but if $f(x) = a^x$ then $f'(x) = a^x ln(a)$
  - if $f(x) = ln(x)$ then $f'(x) = 1/x$
- Rules of derivatives:
  - Product rule: $(fg)' = f'g + g'f$
  - Quotient rule: $\left(\frac{f}{g}\right)' = \frac{f'g - g'f}{g^2}$
  - Chain rule: if $f(x) = h(g(x))$ then $f'(x) = h'(g(x)) \times g'(x)$

Normal Distribution Functions - Recap

- Recall from last week: the PDF of a Normal random variable is
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$
- Behaviour is dictated by two parameters:
  - $\mu \in \mathbb{R}$: mean, represents the centre of the distribution
  - $\sigma > 0$: standard deviation, represents the spread/variability of the distribution
- A special case is the **Standard Normal** distribution, where $\mu = 0$ and $\sigma = 1$, where we sometimes use $Z \sim N(0, 1)$ to distinguish from other Normals

Normal Distribution Functions - Recap

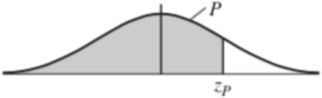
- Recall also that the Normal distribution does not have a closed-form expression for the CDF.
- However, we may consider the CDF of the standard Normal distribution:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

  as an integral over the standard Normal PDF.
- Again this has no closed-form solution, but it has been evaluated using numerical techniques.
- Thus computing probabilities using the CDF $\Phi(z)$ is done through tables of standard Normal values (see Appendix B of the textbook).

Standard Normal CDF Table - Recap

TABLE 2 Cumulative Normal Distribution—Values of P Corresponding to z_p for the Normal Curve



z is the standard normal variable. The value of P for $-z_p$ equals 1 minus the value of P for $+z_p$; for example, the P for -1.62 equals $1 - .9474 = .0526$.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879

Normal Distribution Functions

- Thus, if there were a way to move from some $N(\mu, \sigma^2)$ to the $N(0,1)$, then we would be able to compute probabilities using the standard Normal table.
- Turns out we can!
- We need to be able to **transform** a Normal RV X to a standard Normal RV Z, by manipulating the CDF.
- This involves writing one RV as a function of another RV, and finding the new PDF or CDF.
- These functions can be anything, but let's start with a simple one: $Y = aX + b$, where $a > 0$.

Normal Distribution Functions

- Suppose that $X \sim N(\mu, \sigma^2)$, and we are interested in the CDF for the transformed variable $Y = aX + b$, where $a > 0$.
- We can start by directly writing the definition of the CDF of Y: $F_Y(y) = P(Y \leq y)$
- Since we know the CDF of X, we just need to rewrite $F_Y(y)$ in terms of X:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) \\ = P\left(X \leq \frac{y-b}{a}\right)$$

- But I know what the CDF of X is:

$$F_{\color{red}Y}(y) = P\left(X \leq \frac{y-b}{a}\right) = F_{\color{red}X}\left(\frac{y-b}{a}\right)$$

Normal Distribution Functions

- This means I am able to write my CDF of Y as the CDF of X evaluated at a back-transformed value of Y.
  - this value is just some x that my original X can take, written as a function of y
- We sometimes call this the direct method of transforming variables, because we are just brute forcing it.
- The direct method will work for non-Normal random variables, as well as generally simple functions of X.
- So once I have my new CDF, how do I find the PDF?

Normal Distribution Functions

- As with any generic CDF, to find the PDF you must take the derivative.
- Recall that we have the relationship: $f(x) = \frac{d}{dx}F(x)$
- To find the PDF of Y, we will use the chain rule:

$$f_Y(y) = \frac{d}{dy}F_X\left(\frac{y-b}{a}\right) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right)$$

- So we are able to get a Normal PDF back when we transform a Normal RV (and we haven't even used that X was Normal!):

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y-(b+a\mu)}{a\sigma}\right)^2\right]$$

Generalized Normal Distribution

- So we can see that the form of the PDF of Y is the same as the PDF of X but with a new mean and variance:

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-(b+a\mu)}{a\sigma}\right)^2\right]$$

- We can show that we arrive at the same conclusion if $a < 0$.
- This motivates the notion of the generalized Normal Distribution:

Generalized Normal Distribution

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$

Finding Probabilities of a Normal Distribution

- The intuition of this is that if all we are doing to X is shifting where it is centred and scaling its variance, then we are just getting back a slightly different looking Normal.
- We may use this result now to find probabilities from any Normal distribution.
- Suppose we are interested in finding $P(x_0 < X < x_1)$, for some $X \sim N(\mu, \sigma^2)$
- Since when we transform in this way, we are shifting the mean and scaling the variance, can we shift and scale X so that we end up with Z?

Finding Probabilities of a Normal Distribution

- Consider the random variable

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

- This is a transformation of the form $aX + b$ where $a = {}^{1}/_{\sigma}$ and $b = -{}^{\mu}/_{\sigma}$.
- Therefore, by the previous result,

$$Z \sim N\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2\right) = N(0, 1)$$

and thus

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

Finding Probabilities of a Normal Distribution

- Therefore, to solve our original probability, $P(x_0 < X < x_1)$, we can use the standard Normal CDF by

$$\begin{aligned} P(x_0 < X < x_1) &= F_X(x_1) - F_X(x_0) \\ &= \Phi\left(\frac{x_1 - \mu}{\sigma}\right) - \Phi\left(\frac{x_0 - \mu}{\sigma}\right) \end{aligned}$$

  which now just involves finding those values of $\Phi(x)$ from the standard Normal table.

- Recall that for some $z*$ value, we locate this $z*$ by combining the margins of the table, and trace inwards to get $P(Z \leq z*)$

Example: IQ Scores

IQ scores on a standardized test are approximately Normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 15$. An individual is selected at random. What is the probability that his/her score is between 120 and 130?

- In order to use the standard Normal table, we must transform each of these 2 IQ scores into z's by

$$z_1 = \frac{120 - 100}{15} = 1.33 \quad \text{and } z_2 = \frac{130 - 100}{15} = 2$$

- Now rather than finding the probability with the original IQ score distribution, we use the standard Norma:

$$P(120 < X < 130) = P(1.33 < Z < 2) = \Phi(2) - \Phi(1.33)$$

- We can now look these up in the table

Example: IQ Scores (continued)

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.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
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

Exercise - Give it a try!

Let $X \sim N(\mu, \sigma^2)$. Find the probability that X is less than σ away from μ; that is, find $P(|X - \mu| < \sigma)$.

Direct Method of Transformation

- We can of course use the direct method to find the PDF of RVs that results from transformations other than those of the form $aX + b$.
- Let U be a uniform random variable on $[0, 1]$ and let $V = {}^1\!/_U$.
- We can find the CDF of V by

$$\begin{aligned} F_V(v) &= P(V \le v) = P({}^1\!/_U \le v) \\ &= P(U \ge {}^1\!/_v) \\ &= 1 - P(U \le {}^1\!/_v) \\ &= 1 - {}^1\!/_v \end{aligned}$$

where we get the last line by using the CDF of the Uniform(0,1)

Direct Method of Transformation

- When you do a transformation of this kind (or any kind), it is important to adjust the range of values that V can take
  - U can take values from 0 to 1
  - In order for $F_V(v)$ to be a valid CDF, it cannot take negative values.
  - Therefore, V can only take on values larger than 1 ($v \geq 1$)
  - For $v < 1$, $F_V(v) = 0$.
- Finally to get the PDF, we take the derivative of $F_V(v)$:

$$f_V(v) = \begin{cases} \frac{1}{v^2}. & 1 \leq v < \infty \\ 0, & \text{otherwise} \end{cases}$$

Outline

Chi-Square Random Variable

- The Normal distribution plays a crucial role in statistics.
- In particular, a number of other distributions can be derived from a transformation of a standard Normal RV.
- One such distribution is the **Chi-Square** distribution.
- It is very important to remember how various distributions are related to each other.

Chi-Square Distribution

- Suppose $Z \sim N(0,1)$ and we apply the transformation $X = Z^2$.
- We can still use the direct method of transformations to find the CDF of X
- We get

$$
\begin{aligned}
F_X(x) &= P(X \le x) = P(Z^2 \le x) \\
&= P(-\sqrt{x} \le Z \le \sqrt{x}) \\
&= P(Z \le \sqrt{x}) - P(Z \le -\sqrt{x}) \\
&= \Phi(\sqrt{x}) - \Phi(-\sqrt{x})
\end{aligned}
$$

Chi-Square Distribution

- To get the PDF of X, we take the derivative because $\Phi^{'}(x) = \phi(x)$
- Again, we need to use the chain rule:

$$\Phi^{'}(\sqrt{x}) = \phi(\sqrt{x})\frac{d}{dx}\left(\sqrt{x}\right)$$

- This will give us the PDF

$$f_X(x) = \frac{1}{2}x^{-1/2}\phi\left(\sqrt{x}\right) + \frac{1}{2}x^{-1/2}\phi\left(-\sqrt{x}\right)$$

- We can actually simplify this further when we realize that $\phi(x)$ is a symmetric function, i.e. $\phi(x) = \phi(-x)$ because Normals are symmetric:

$$f_X(x) = x^{-1/2}\phi(\sqrt{x})$$

Chi-Square Distribution

- From here we can replace ϕ with the density of the standard Normal, replacing x with $\sqrt{x}$

$$f_X(x) = \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-x/2}, \; x \geq 0$$

- This is the PDF for a Chi-Square random variable.
- It is worth noting that the above density looks quite similar to

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

(Gamma PDF) if $\alpha = \lambda = 1/2$, since $\Gamma(1/2) = \sqrt{\pi}$

- So the Chi-Square here is the same as a $\text{Gamma}(0.5, 0.5)$

Outline

Transformations thus far

- Up until now, we have been using the same procedure to find the PDF of a transformed variable
  - Find the CDF of the transformed variable $Y = g(X)$ by substituting $X = g^{-1}(Y)$ into the CDF of X
  - Next differentiate the CDF to find the PDF of Y, which often involves also differentiating $g^{-1}(X)$
- This procedure can be used to prove a more general method of transformation: the Monotone Transformation Method

Monotone Transformation Method

Monotone Transformation Method

Let X be a continuous random variable with density $f(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose that $f(x) = 0$ is x is not in I. Then Y has the density function

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

for y such that $y = g(x)$ for some x, and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I. Here g^{-1} is the inverse function of g; that is, $g^{-1}(y) = x$ if $y = g(x)$.

- This is just a concise expression for the procedure we have been using up to now
- You may use whichever one is easier for you.

Example: Sugar Production and Profit

Suppose X is a random variable representing the amount of refined sugar (in tonnes per day) produced in some process, with density function

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Now consider $Y = 3X - 1$ representing the daily profit. Find the density function of Y.

- We have that our transformation function is $y = g(x) = 3x - 1$.
- Thus the inverse of this function is $x = g^{-1}(y) = \frac{y+1}{3}$
- To use the monotone transformation method, we also need to differentiate $g^{-1}(y)$:

$$\left| \frac{d}{dy} \left(\frac{y+1}{3} \right) \right| = \frac{1}{3}$$

Example: Sugar Production and Profit (cont.)

- We can now plug in everything to the monotone transformation expression:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$
$$= 2\left(\frac{y+1}{3}\right) \times \frac{1}{3} = \frac{2(y+1)}{9}$$

- Finally we need to specify for what values of y the function is defined over:

$$0 \le x \le 1 \Rightarrow 0 \le \frac{y+1}{3} \le 1 \Rightarrow -1 \le y \le 2$$

- Therefore the PDF of Y is

$$f_Y(y) = \begin{cases} \frac{2(y+1)}{9}, & -1 \le y \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Exercise - Give it a try!

Let X have probability density function given by

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the density function of $Y = -4X + 3$.

Monotone Tranformations

- The key element of the Monotone Transformation Method is that it can only be used when your transformation is a monotone function
- However, since it only needs to be monotone in an interval I, you can still use this method for non-monotone functions
- Requires you to divide the domain of the function into intervals in which the function is monotone
- Then just apply the monotone transformation method to each sub-interval of the domain.

Example: Non-monotonic transformation

For some X with continuous PDF, find the density function for the transformation $Y = X^2$ for $-\infty < x < \infty$.

- Obviously this transformation is not monotone.
- Let's see how we would approach this using the direct method:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

- To get the PDF, we would just differentiate both terms:

$$\begin{aligned} f_Y(y) &= F_X^{'}(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - F_X^{'}(-\sqrt{y})\left(\frac{-1}{2\sqrt{y}}\right) \\ &= \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right],\ \ y > 0 \end{aligned}$$

Example: Non-monotonic transformation (cont.)

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right],\ y > 0$$

- So now $f_Y(y)$ is the sum of two pieces, where each piece is defined on a subinterval that makes $g(X)$ a monotonic function on that interval.
  - $f_X(\sqrt{y})$ is defined for $x \geq 0$
  - $f_X(-\sqrt{y})$ is defined on $x < 0$
- Even though we used the direct method to find the PDF of Y, we could have instead used the monotone transformation method two separate times:
  - once for $x \geq 0$ resulting in $\frac{1}{2\sqrt{y}} f_X(\sqrt{y})$
  - once on $x < 0$ resulting in $\frac{1}{2\sqrt{y}} f_X(-\sqrt{y})$
  - then just add them together.

Outline

Transformations with the Uniform Distribution

- We now will present two results based off of transformations involving the Uniform distribution
- These are useful when attempting to generate pseudorandom numbers from a particular distribution
  - "pseudo" because not completely randomly generated
  - we use an algorithm or rule to generate them
- Recall: a Uniform random variable defined on an interval $[a, b]$ has PDF

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}$$

and CDF

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases}$$

Probability Integral Transformation

Probability Integral Transformation

Let $Z = F_X(X)$, then Z has a Uniform distribution on $[0, 1]$.

- This means that the transformation we are making on X is to apply the function corresponding to the CDF of X on X.

Proof:

$\square$

Example: Transformation with Exponential CDF

Suppose $X \sim Exp(\lambda)$ and we wish to find the distribution of $Z = F_X(X)$, i.e. my transformation function is $g(X) = 1 - e^{-\lambda X}$.

- We can show this in the same way as the proof, using the direct method:

Inverse Integral Transformation

Inverse Integral Transformation

Let U be uniform on $[0, 1]$, and let $X = F^{-1}(U)$. Then the CDF of X is F.

- This is essentially just the reverse of the probability integral transformation.

Proof:

$\square$

Example: Generating from Exponential Distribution

- Suppose I want to generate values from an Exponential distribution.
- I can do this if I first generate values from a Uniform$[0, 1]$ and then apply a transformation.
- To get an Exponential, my transformation must be the inverse of my Exponential CDF:

$$\begin{aligned} u &= 1 - e^{-\lambda x} = F_X(x) \\ e^{-\lambda x} &= 1 - u \\ -\lambda x &= \log(1 - u) \\ x &= -\,{}^{\log(1-u)}/_{\lambda} = F^{-1}(u) \end{aligned}$$

- So if u are values from a Uniform$[0, 1]$, then under this transformation, I can generate values from an Exponential

Remarks

- The probability integral transform and the inverse integral transform arise as direct results of the monotone transformation method/direct method
- In most cases, you will only need to concern yourself with the direct method and the monotone transformation method
- The two integral transformations are really only relevant when dealing with Uniform distributions, or specifically when the transformation function is a CDF.