Stats 120B: Jan, 09&11,2017

- Prof: Weining Shen (weinings, 2204 Bren Hall)
- TA: Fan Yin (yinf2)
- Reader: Wei Hu (huw5)

About this class

- All the materials are on the class website.
- One midterm, one final
- HW: once a week, due on Wed. (HW number)
- Office hour: Mon (2-3pm) /Tue (?)/Thurs (?)
- Fan Yin (3:30 - 5:30pm)
- Slides most of the time (posted on the website)
- Hard homework, easy exams.

- Text: book by DeGroot and Schervish.
- Some other stat classes: 120 series, 110, 111, 7
- Software: R
- Rule 1: NO CHEATING
- Rule 2: ASK QUESTIONS!!!

Topics this week

- Review: 120A
- Gamma distribution (Ch 5.7)
- Beta distribution (Ch 5.8)
- A review on moment generating function (Ch 4.4)
- Function of random variables (Ch 3.8 & 3.9)

Stats 120A: Probability

- The world is full of randomness...
- You ask someone out for a date. Y or N?
- Basketball game: Clippers vs. warriors
- Binary random variable... what distribution?
- Soccer game: Win/lose/draw? (Still binary?)
- When's the next major earthquake in SoCal?
- Random variables? Discrete/Continuous?

Discrete r.v.

- Example: A pocket, 10 balls, 9 black, 1 red

- (i) Take a ball, what's the probability of getting a red?

  One red ball, ten balls in total, so the prob is $1/10$

- (ii) Take a ball, find that it is a black one, throw it away, take a second ball, what's the prob of getting a red?

  One red ball, nine balls in total, so the prob is $1/9$

- (iii) Take a ball, don't look at it/throw it away, take a second ball, what's the prob of getting a red?

  P(2nd is red) = P(2nd is red | 1st is black) P(1st is black) + P(2nd is red | 1st is red) P(1st is red)

  $$= 1/9 \times 9/10 + 0 \times 1/10 = 1/10$$

Example cont.

- (iv) Take a ball, it's a red one, put it back, take a second ball, what's the prob of red?

  One red ball, ten balls in total, so the prob is still $1/10$.
  Note that two events are independent.

- (i) Probability
- (ii) Conditional prob.
- (iii) law of total probability
- (iv) Independence; sample w replacement.

Discrete distributions

- OK... so I have a discrete random variable X.
- That means X takes a finite number of values with probs. How do I describe it?
- Examples: uniform, Bernoulli, Binomial, geometric, Poisson etc.

- For any distribution, you should look at its —
- (i) Support, shape, symmetry, range, skewness, tail
- (ii) Mean, variance, moments, mode, quantiles
- (iii) Density function, distribution function, pmf, mgf etc.
- Example: ask someone out? support: Y/N.
- Yelp score? support: [0,5].
- Quantile? Median, 25%-quantile etc.?

Example: binomial distribution $\text{Bin}(n, p)$

- What's its support?
- $0, 1, \ldots, n$
- Mean/Variance?
- $np, np(1-p)$
- When's it symmetric?
- $p = 1/2$
- What about Poisson distribution $\text{Poisson}(\lambda)$?

Continuous distributions

- Continuous r.v. X. What's the big difference?
- That means X takes <u>a continuous range of values</u> with probs.
- Examples: Uniform, Normal, Exponential, Gamma, Beta etc.
- More examples: Laplace, Cauchy, Weibull, Pareto, Extreme value, log-normal etc.

- Wait.. why do I need to learn all those different distributions?
- What is the problem of using normal distribution?

(a) It requires the data being symmetric;

(b) It takes value on the entire real line, which is not appropriate to model positive random variables, e.g., income, time etc.

(c) The tail probability is too small, may not be able to model rare-events, such as earthquakes.

- What is the problem with exponential distribution?

  (a) Its mode is at 0. Can't use it to model data such as income, age etc.

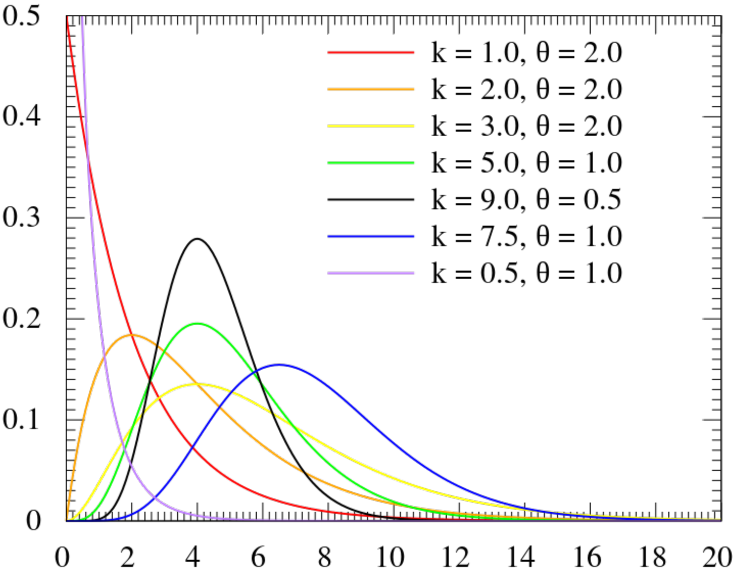
  (b) Say X $\sim$Exp(β), then its expectation is $1/\beta$ and its variance is $1/\beta^2$. In practice, it's hard to find data satisfy this requirement, i.e., mean^2 $=$ variance.

- What is the problem with uniform distribution?

  The density has the same value throughout, there is no variation.

  The worlds is fair? hahaha..

Gamma distribution (Ch 5.7)



Gamma distribution

- What does gamma distribution offer?
  (a) It provides a variety of shapes
  (b) It take values only on positive real line.
  (c) It includes exponential distribution as a special example.

- Definition Let $\alpha, \beta > 0$, a random variable X follows a *gamma distribution* with parameters α and β if the pdf of X is

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \; x > 0.$$

- Definition For any $\alpha > 0$, define its *Gamma function* by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

By this definition, gamma function is always positive.

Theorem 5.7.1 For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
Proof

$$
\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx \\
&= \int_0^\infty x^{\alpha-1} d(-e^{-x}) \qquad * \int u dv = uv - \int v du. \\
&= -x^{\alpha-1} e^{-x} |_0^\infty + \int_0^\infty e^{-x} d(x^{\alpha-1}) \\
&= 0 + \int_0^\infty e^{-x} (\alpha - 1) x^{\alpha-2} dx \\
&= (\alpha - 1) \int_0^\infty e^{-x} x^{\alpha-2} dx \;\; \text{(by definition, the integral is } \Gamma(\alpha-1)\text{)} \\
&= (\alpha - 1)\Gamma(\alpha - 1)
\end{aligned}
$$

$$\Gamma(1) = \int_0^\infty x^{1-1} e^{-x} dx = 1$$

.

If n is an integer, then $\Gamma(n) = (n-1)!$. Here's the proof.

$$\begin{aligned}
\Gamma(n) &= (n-1)\Gamma(n-1) \\
&= (n-1)(n-2)\Gamma(n-2) \\
&\cdots \\
&= (n-1) \times (n-2) \times \cdots 2 \times 1 \times \Gamma(1) \\
&= (n-1)!
\end{aligned}$$

- Given the pdf $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \; x > 0.$

- Verify it is a valid density function

(a) $f(x) \geq 0$ for every $x > 0$
(b) We need to show that $\int_0^\infty f(x) = 1.$ This is true because

$$\int_0^\infty f(x) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx$$

Change of variable, let $y = \beta x, dy = \beta dx$

$$= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{y}{\beta}\right)^{\alpha-1} e^{-y} \frac{dy}{\beta}$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} dy$$

By definition, this integral is $\Gamma(\alpha)$

$$= 1.$$

- Expectation is

$$\begin{aligned}
\mathsf{E}X &= \int_0^\infty x f(x) dx \\
&= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\beta x} dx \quad \text{*Change of variable, let } y = \beta x, dy = \beta \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\beta}\right)^\alpha e^{-y} \frac{dy}{\beta} \\
&= \frac{1}{\beta\Gamma(\alpha)} \int_0^\infty y^\alpha e^{-y} dy \qquad \text{Definition, this integral is } \Gamma(\alpha+1) \\
&= \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} \\
&= \frac{\alpha}{\beta}.
\end{aligned}$$

How to get the variance?

$$
\begin{aligned}
\mathsf{E}X^2 &= \int_0^\infty x^2 f(x)dx \\
&= \int_0^\infty x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\beta x} dx \qquad \text{let } y = \beta x \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\beta}\right)^{\alpha+1} e^{-y} \frac{dy}{\beta} \\
&= \frac{1}{\beta^2 \Gamma(\alpha)} \int_0^\infty y^{\alpha+1} e^{-y} dy \qquad \text{Definition, this integral is } \Gamma(\alpha+2) \\
&= \frac{\Gamma(\alpha+2)}{\beta^2 \Gamma(\alpha)} \\
&= \frac{\alpha(\alpha+1)}{\beta^2}.
\end{aligned}
$$

Variance is

$$
\begin{aligned}
\mathsf{Var} X &= \mathsf{E} X^2 - (\mathsf{E} X)^2 \\
&= \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 \\
&= \frac{\alpha}{\beta^2}.
\end{aligned}
$$

Connection with exponential distribution

- Exp(β) is a special case of Gamma(α, β) with $\alpha = 1$.
  Look at the density function of Gamma(α, β):

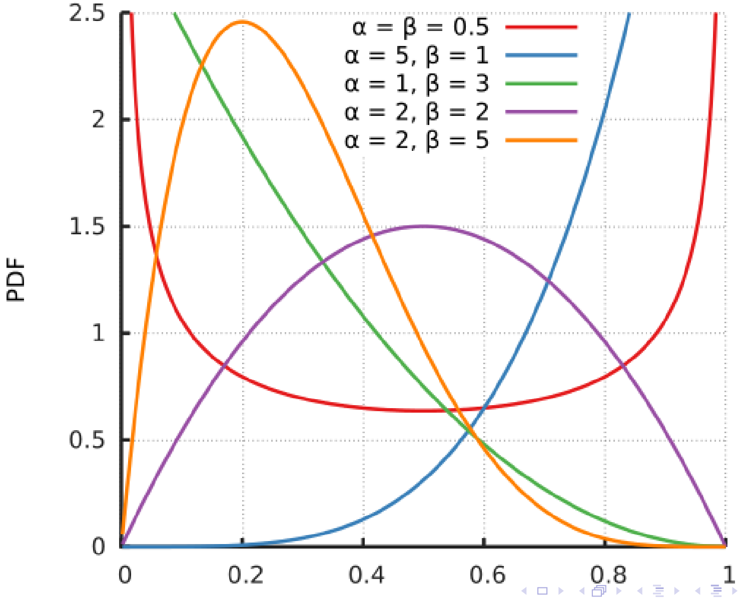
  $$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

  Let $\alpha = 1$, then it becomes $\beta e^{-\beta x}$ ($\Gamma(1) = 1$), which is the pdf of Exp(β).
- If $X_i \sim$ Gamma(α_i, β) independently for $i = 1, \dots, N$, then

  $$\sum X_i \sim \text{Gamma}\left(\sum_{i=1}^{N} \alpha_i, \beta\right).$$

  Proof... Homework
  Hint: use MGF

Beta distribution (Chap 5.8)



Beta distribution (Chap 5.8)

What does Beta distribution offer?

- (a) It is defined on $(0, 1)$, ideal for modeling variables such as proportion, response rate, probability etc.
- (b) A variety of shapes, flexible.
- (c) Includes uniform distribution on (0,1) as a special example.

- Definition Let $\alpha, \beta > 0$, a random variable X follows a *beta distribution* with parameters α and β if the pdf of X is

$$f(x) = \frac{1}{\mathsf{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \ x \in (0,1).$$

- Definition For any $\alpha, \beta > 0$, define their *beta function* by

$$\mathsf{B}(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

- Beta function is always positive.

Given $f(x) = \frac{1}{\mathsf{B}(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)},\ x \in (0,1).$

Verify it is a valid density function

(a) $f(x)$ is positive for $x \in (0,1)$.

(b) Would like to show $\int_0^1 f(x)dx = 1$.

$$\begin{aligned}
\int_0^1 f(x)dx &= \int_0^1 \frac{1}{\mathsf{B}(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)}dx \\
&= \frac{1}{\mathsf{B}(\alpha,\beta)} \underline{\int_0^1 x^{(\alpha-1)}(1-x)^{(\beta-1)}dx} \\
&\quad **\text{by the definition of beta function, this integral is } \mathsf{B}(\alpha,\beta) \\
&= 1.
\end{aligned}$$

Recall $f(x) = \frac{1}{\mathsf{B}(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)},\ x \in (0,1)$.

$$
\begin{aligned}
\mathsf{E}X &= \int_0^1 x f(x) dx \\
&= \int_0^1 x \frac{1}{\mathsf{B}(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)} dx \\
&= \frac{1}{\mathsf{B}(\alpha,\beta)} \underline{\int_0^1 x^{\alpha}(1-x)^{(\beta-1)} dx} \quad ** \text{ this integral is } \mathsf{B}(\alpha+1,\beta) \\
&= \frac{\mathsf{B}(\alpha+1,\beta)}{\mathsf{B}(\alpha,\beta)} \qquad ** \text{ Recall } \mathsf{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \quad * \text{ Recall } \Gamma(\alpha+1) = \alpha\Gamma(\alpha),\ \forall \alpha > 0 \\
&= \frac{\alpha}{\alpha+\beta}.
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{E}X^2 &= \int_0^1 x^2 f(x)dx \\
&= \int_0^1 x^2 \frac{1}{\mathsf{B}(\alpha,\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)}dx \\
&= \frac{1}{\mathsf{B}(\alpha,\beta)} \underline{\int_0^1 x^{\alpha+1}(1-x)^{(\beta-1)}dx} \quad ** \text{ this integral is } \mathsf{B}(\alpha+2,\beta) \\
&= \frac{\mathsf{B}(\alpha+2,\beta)}{\mathsf{B}(\alpha,\beta)} \qquad ** \text{ Recall } \mathsf{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+2)} \quad * \text{ Recall } \Gamma(\alpha+1) = \alpha\Gamma(\alpha),\ \forall\alpha > 0 \\
&= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}
\end{aligned}
$$

So the variance of Beta(α, β) is

$$\mathsf{E}X^2 - (\mathsf{E}X)^2 = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2$$
$$= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

How to find the mode for Beta dist?

For any pdf $f(x)$, we want to find x that maximizes $f(x)$, this x is called the mode. Here are the general steps you could use:

- (a) Drop the constant in $f(x)$, i.e., terms that doesn't involve x.

$$\frac{1}{\mathrm{B}(\alpha, \beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)} = C x^{(\alpha-1)}(1-x)^{(\beta-1)}$$

- (b) Take the log

$$C x^{(\alpha-1)}(1-x)^{(\beta-1)} \to \log C + (\alpha-1)\log x + (\beta-1)\log(1-x)$$

- (c) Take the derivate.
- (d) Set the derivative to 0, solve x, the solution is the mode.

A review on moment generating function (mgf)

- mgf uniquely determines the distribution.

  Say a random variable is a person, then the distribution function is his/her SSN, density function is his/her driver license.

- Similarly, mgf is the passport.

- If you can find any of these functions, you will be able to identify the random variable, such as the name of the dist and the value of parameters.

When you may consider using mgf?

- Example 1: $X_1, \ldots, X_k$ independently follow some distributions. Determine the distribution of $X_1 + \cdots + X_k$.
- Example 2: X follows some distribution, determine the distribution of cX.
- Example 3: X follows some distribution, determine the distribution of $X + b$.
- In general, consider mgf if you know $X, Y, Z...$ follow some distribution independently, and you want to find the distribution of linear combination of $X, Y, Z...$

More on mgf

- Definition Let X be a r.v. For each real number t, define

$$M_X(t) = E(e^{tX})$$

- mgf is a function of t.
- mgf does NOT exist for some values of t.
- You can obtain moments of X by taking derivatives on its mgf. This is where the name "moment generating" comes from...
- Mgf does not exist for certain distributions
- A more general definition, called characteristic function

Useful facts about mgf

- Theorem 4.4.4 Suppose that $X_1, \ldots, X_n$ are independent r.v.s with mgf $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $Y = \sum_{i=1}^n X_i$, its mgf is

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t).$$

- Theorem 4.4.3 Let X be a r.v. with mgf $M_X(t)$. Define $Y = aX + b$, then its mgf is

$$M_Y(t) = e^{bt} M_X(at).$$

An example

- Let $X \sim N(\mu, \sigma^2)$. Its mgf is $M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}$.
- Suppose that $W \sim N(65, 1)$ and $V \sim N(68, 3)$, independently.
- Let $X = V - W$. Find the distribution of X using mgf.
- $M_X(t) = M_V(t) \times M_{-W}(t)$.
- $M_V(t) = \exp\{68t + \frac{3}{2}t^2\}$
- $M_{-W}(t) = M_W(-t) = \exp\{-65t + \frac{1}{2}t^2\}$
- So $M_X(t) = M_V(t) \times M_{-W}(t) = \exp\{3t + 2t^2\}$.
- $M_X(t)$ looks like the mgf of normal distribution... with $\mu = 3$ and $\sigma^2 = 4$.
- So $X \sim N(3, 4)$.

Function of random variables (Chap 3.8 & 3.9)

- (1) Given the distribution of X, find the distribution of $r(X)$.
- (2) Given the distribution of $X_1, \ldots, X_n$, find the distribution of minimum (or maximum) of $X_1, \ldots, X_n$.

  e.g., Extreme cases, the lowest winter temperature, quality control (the worst product should still meet high standard).
- For (1), the answer depends on what type of distribution (discrete or continuous) X has.

Function of discrete r.v.

X	-1	0	1	2
$f(X)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

What's the distribution of $Y = X^2$?

Here are the steps:

(1) Find all the possible values that Y could take:

Y could take $\{-1^2, 0^2, 1^2, 2^2\} = \{0, 1, 4\}$.

(2) For each value y of Y, find x such that $x^2 = y$, then combine their probabilities together.

$Y = X^2$	0	1	4
X	0	-1, 1	2
$f(Y)$	$\frac{1}{4}$	$\frac{1}{4} + \frac{1}{4}$	$\frac{1}{4}$

Theorem 3.8.1 Let X have a discrete distribution with probability mass function (pmf) f, and let $Y = r(X)$ for some function of r defined on the set of possible values of X. For each possible value y of Y, the pmf g of Y is

$$g(y) = \mathsf{Pr}(Y = y) = \mathsf{Pr}[r(X) = y] = \sum_{x:r(x)=y} f(x).$$

Function of a continuous r.v.

Say $X \sim \text{Exp}(1)$, what's the distribution of $Y = 1 - e^{-X}$?

(1) Find the range of Y

X is on $(0, \infty)$, so Y is from $1 - e^{-0}$ to $1 - e^{-\infty}$, i.e., $(0, 1)$.

(2) Write the distribution function of Y, and express Y using X

$$
\begin{aligned}
P(Y \leq y) &= P(1 - e^{-X} \leq y) \\
&= P(e^{-X} \geq 1 - y) \\
&= P(-X \geq \log(1 - y)) \\
&= P(X \leq -\log(1 - y)) \quad * \text{recall CDF of Exp}(1) \text{ is } 1 - e^{-x} \\
&= 1 - e^{\log(1-y)} = y, \quad ***\text{Uniform distribution.}
\end{aligned}
$$

(3) That's the CDF of Y, u can take the derivative to obtain pdf.

Distribution of max/min

Suppose that $X_1, \ldots, X_n \sim F(x)$ i.i.d. Define

$$Y_n = \max(X_1, \ldots, X_n); \quad Y_1 = \min(X_1, \ldots, X_n).$$

What's the distribution of Y_1 and Y_n?

(1) Max is easier...

$$\begin{aligned} P(Y_n \leq y) &= P(\max(X_1, \ldots, X_n) \leq y) \\ &= P(X_1 \leq y, \ldots, X_n \leq y) \ * \text{max} \leq y \text{ means every } X \leq y. \\ &= F(y)^n. \end{aligned}$$

(2) Min is tricky...

$$P(Y_1 \le y) = P(\min(X_1, \ldots, X_n) \le y)$$

$*$ this doesn't imply anything useful.

$$
\begin{aligned}
P(Y_1 \le y) &= 1 - P(Y_1 > y) \\
&= 1 - P(\min(X_1, \ldots, X_n) > y) \quad **\text{Now it's useful} \\
&= 1 - P(X_1 > y, \ldots, X_n > y) \;\; * \text{ min} > y \text{ means evy } X > y. \\
&= 1 - \{P(X_1 > y)\}^n \\
&= 1 - \{1 - F(y)\}^n.
\end{aligned}
$$

Example

Suppose that $X_1, \ldots, X_n$ are i.i.d random samples following unif$(0, 1)$. Find the distribution of their max Y_n and min Y_1.

(1) The distribution function of X is $F(x) = x$.

(2) $P(Y_n \leq y) = F(y)^n = y^n$, the pdf is ny^{n-1}.

(3) $P(Y_1 \leq y) = 1 - \{1 - F(y)\}^n = 1 - (1 - y)^n$, the pdf is $n(1 - y)^{n-1}$.

(4) Don't forget the range/support: $y \in (0, 1)$.

Summary

- Discrete/continuous distributions, the concept of support.
- View a distribution from different aspects
- How to get the mode/median?
- Moment generating function: what's it? when to use it? how to use it?
- Function of random variables: min/max, transformation (discrete,continuous).