F19 STA 100 A01 Discussion 06

Yishan Huang

2019/11/05

Discussion Time: Tuesday 8:00 – 8:50 am, Haring Hall 1204.

Notes: https://github.com/Hahahuo-13316/sta100-a01-fall19

Office hour: Tuesday 12:00 – 1:00 pm, Mathematical Sciences Building 1117.

Email: yishuang@ucdavis.edu

Quiz: This Friday.

Continuous Random Variables (e.g, normal distribution, uniform distribution, etc)

c.d.f vs. p.d.f:

- The c.d.f of a random variable X is $F(x) = P(X \leq x)$. It is the cumulative probability for X from $-\infty$ to x. For a continuous random variable X, F is always continuous.
- The p.d.f of a continuous random variable X is $f(x)$. It satisfies that $F(x + \varepsilon/2) - F(x - \varepsilon/2) = P(x - \varepsilon/2 < X < x + \varepsilon/2) = \varepsilon \cdot f(x)$. For common continuous random variables, we always have F is differentiable and

$$F'(x) = f(x), \quad \int_a^b f(x)\mathrm{d}x = F(b) - F(a) = P(a < X < b).$$

  That is to say, the probability that X lies in (a, b) is the area under the curve $f(x)$ between $x = a$ and $x = b$.
  - Notice that, $f(x)$ is **NOT** the probability of $P(X = x)$. In fact, for continuous random variable X, $P(X = x) = 0$ for any x. X may have some probability to be close to x, but it cannot exactly equal to x.

Facts of normal distribution

	Standard Normal	Normal	Uniform
Form	$Z \sim N(0,1)$	$X = \sigma Z + \mu \sim N(\mu, \sigma^2)$	$X \sim \text{Unif}([a,b])$
Mean	0	μ	$\frac{a+b}{2}$
Variance	1	σ^2	$\frac{(b-a)^2}{12}$
cdf	$P(Z \leq z) = F(z)$	$P(X \leq x) = F\left(\frac{x-\mu}{\sigma}\right)$	$\frac{x-a}{b-a}$ for $a \leq x \leq b$
pdf	$f(z)$	$f_X(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$	$\frac{1}{b-a}$ for $a \leq x \leq b$
Quantile with probability p	q where $F(q) = p$	$\sigma q + \mu$ where $F(q) = p$	$q = pb + (1-p)a$

Table 1: comparing continuous random variables

Remarks:

- We can calculate $F(z)$ by referring to the normal table. Because we can always know $P(0 < Z < z_1)$ for any $z_1 > 0$. Then, we can derive by $F(z) = \frac{1}{2} + P(0 < Z < z)$ for $z > 0$; and $F(z) = \frac{1}{2} - P(0 < Z < (-z))$ for $z < 0$.

- Similarly, for finding quantiles, if $p > 1/2$, we can find it with $P(0 < Z < q) = p - \frac{1}{2}$; if $p < 1/2$, we can find it with $P(0 < Z < (-q)) = \frac{1}{2} - p$.
- In uniform distribution, the range is essential. Hence we should write

$$f_U(u) = \begin{cases} \frac{1}{b-a}, & a \le u \le b; \\ 0, & \text{otherwise.} \end{cases}$$

Sampling distributions

- When we draw a sample of size n: $\{Y_1, \dots, Y_n\}$ from a population P, then the sampling distribution of the sample mean is the distribution of $\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$.
- If the population P has mean μ and variance σ^2, then $\bar{Y}$ has mean μ and variance σ^2/n and hence standard deviation $\sigma/\sqrt{n}$.

Problems

- If a population follows Bernoulli(p) distribution, i.e, with the distribution

x	0	1
$P(X = x)$	$1 - p$	p

  then what is sampling destribution of the sample mean, for a sample with size n?
- In a standard test, the score of the population (which is very large) has normal distribution $N(60, 10^2)$. Here, we suppose the score can take any real number.
  - If we pick a random student from the population and let X represents his/her score. What is the probability of $55 < X < 75$?
  - If a student do not want to be either the top 5% student or the bottom 20% student, then what is the possible range of the student's score?