STAT 230 Chapter 8 Continuous Probability Distributions

8.1 General Terminology and Notation

cdf: properties
- $\lim_{x\to -\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
- $F(x)$ is a non-decreasing function of x
- $P(a < X < b) = F(b) - F(a)$

$0 = P(X = a) = \lim_{x\to 0} F(a) - F(a-x)$

$$\boxed{P(a<X<b) = P(a\le X\le b) = P(a\le X<b) = P(a<X\le b) = F(b)-F(a)}$$

probability density function : $P(x \le X \le x+\Delta x) = F(X+\Delta x) - F(X)$

defn: The pdf $f(x)$ for a continuous r.v. X is the derivative

$$\boxed{f(x) = \frac{dF(x)}{dx}}$$

properties of pdf
1. $P(a\le X\le b) = F(b) - F(a) = \int_a^b f(x)\,dx$
2. $f(x) \ge 0$ Since $F(x)$ is non-decreasing, the derivative is non-negative.
3. $\int_{-\infty}^{\infty} f(x)\,dx = \int_{\text{all } x} f(x)\,dx = 1$
4. $F(x) = \int_{\infty}^{x} f(u)\,du$

note: for small Δx:
$P(x - \frac{\Delta x}{2} \le X \le x + \frac{\Delta x}{2}) = F(x+\frac{\Delta x}{2}) - F(x - \frac{\Delta x}{2}) \approx f(x)\Delta x$

ex. Consider the spinner example, where

$$F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{4} & 0 < x \le 4 \\ 1 & x > 4 \end{cases}$$

Thus pdf $= f(x) = F'(x) = \frac{1}{4}$ if $0 < x \le 4$, or 0 otherwise.

ex. let X be a continuous random variable with pdf

$$f(x) = \begin{cases} kx^2 & 0<x<1 \\ k(2-x) & 1<x<2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) the constant k

(b) the cumulative distribution function $F(x) = P(X \le x)$

(c) $P(0.5 < X < 1.5)$

Sol. (a)

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_0^1 kx^2dx + \int_1^2 k(2-x)dx + \int_2^{\infty} 0dx$$

$$= k\frac{x^3}{3}\Big|_0^1 + k\left(2x - \frac{x^2}{2}\right)\Big|_1^2$$

$$1 = \frac{5k}{6}$$

$$\therefore k = \frac{6}{5}.$$

(b) $F(x) = P(X \le x) = \int_{-\infty}^{x} f(z)dz = 0 + \int_0^x \frac{6}{5}z^2dz = \frac{6}{5}\frac{z^3}{3}\Big|_0^x = \frac{2x^3}{5}$ if $0<x<1$

$F(x) = P(X \le x) = \int_0^1 \frac{6}{5}z^2dx + \int_1^x \frac{6}{5}(2-z)dz = \frac{12x - 3x^2 - 7}{5}$ if $1<x<2$

$$F(x) = P(X \le x) \begin{cases} 0 & x \le 0 \\ \frac{2x^3}{5} & 0 < x \le 1 \\ \frac{12x-3x^2-7}{5} & 1<x<2 \\ 1 & x \ge 2 \end{cases}$$

(c) $P(0.5 < X < 1.5) = \int_{0.5}^{1.5} f(x)dx = F(1.5) - F(0.5)$

$$= \frac{12(1.5) - 3(1.5)^2 - 7}{5} - \frac{2(0.5)^3}{5} = 0.8.$$

defn Quantiles and Precentiles: Suppose X is a cts r.v. with cdf $F(x)$ the p^{th} quantile of X is the value $q(p)$ such that $P[X \le q(p)] = p$.
The value $q(p)$ is also called the $100p$th percentile of the distri.
If $p = 0.5$ then $m = q(0.5)$ is called the median of X or median of distri.

ex. for the example above, find:

(a) the 0.4 quantile (40^{th} percentile) of the distribution

(b) the median of the distribution.

soln (a) Since $F(1) = 0.4$, the 0.4 quantile is equal to 1.

(b) 0.5 quantile or 50^{th} percentile: $0.5 = F(x) = \frac{12x - 3x^2 - 7}{5}$ for $1 < x < 2$

this yields $x = 1.0235$

defined variables or change of variable.

pdf $\Leftrightarrow$ cdf:

1) Write the cdf of Y as a function of X
2) Use $F_X(x)$ to find $F_Y(y)$. Then if you want pdf $f_Y(y)$, you can differentiate the expression for $F_Y(y)$
3) find the range of values of y

ex. in the earlier spinner example

$$f(x) \begin{cases} 1/4 & 0 < x \le 4 \\ 0 & \text{otherwise.} \end{cases}$$

and

$$F(x) \begin{cases} 0 & x \le 0. \\ x/4 & 0 < x < 4 \\ 1 & x \ge 4 \end{cases}$$

find pdf $Y = X^{-1}$

soln $F_Y(y) = P(Y \le y) = P(X^{-1} \le y)$ step 1

$= P(X \ge y^{-1}) = 1 - P(X < y^{-1})$

$= 1 - F_X(y^{-1})$

$F_Y(y) = 1 - \frac{y^{-1}}{4} = 1 - \frac{1}{4y}$ step2

$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{4y^2}$ for $y \ge \frac{1}{4}$

defn When X is a cts rv we define.

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

ex for the earlier spinner example.

$$f(x) = \begin{cases} 1/4 & 0 < x < 4 \\ 0 & \text{otherwise.} \end{cases}$$

therefore

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = 0 + \int_0^4 x \tfrac{1}{4}\, dx + 0$$

$$= \tfrac{1}{4}\left(\tfrac{x^2}{2}\right)\Big|_0^4 = 2.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^4 x^2 \tfrac{1}{4}\, dx = \tfrac{16}{3}$$

$$\sigma^2 = Var(X) = E(X^2) - \mu^2 = \tfrac{4}{3}$$

8.2 Continuous Uniform Distribution

setup Suppose X takes values in some interval $[a, b]$ with all subintervals of a fixed length being equally likely.
Then X has a cts uniform dis $X \sim U(a, b)$

illustrations: · in the spinner example $X \sim U(0, 4)$
· computers can generate a random number X which appears as though it is drawn from the distribution $U(0,1)$

pdf & cdf

$$\boxed{f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}}$$

$$\boxed{F(x) = \begin{cases} 0 & x < a \\ \int_a^x \frac{1}{b-a}\, dx = \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b. \end{cases}}$$

Mean & Var

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\boxed{Var(X) = \sigma^2 = E(X^2) - \mu^2 = \frac{(b-a)^2}{12}}$$

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

ex. Suppose X has the continuous prob. func

$$f(x) = 0.1e^{-0.1x} \quad \text{for } x>0$$

We will show that the new rv $Y = e^{-0.1X}$ has $U(0,1)$

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(e^{-0.1X} \le y) \\ &= P(X \ge -10\ln(y)) \\ &= 1 - P(X < -10\ln(y)) \\ &= 1 - F_X(-10\ln(y)) \end{aligned}$$

$$F(x) = \int_0^x f(x)dx = \int_0^x 0.1e^{-0.1x}dx = 1-e^{-0.1x}$$

Then $F_Y(y) = 1 - F_X(-10\ln y) = 1-(1-e^{-0.1x})$ $\quad 0<y<1$

$$= e^{-0.1(-10\ln y)} = e^{\ln y} = y$$

$\therefore f(y) = \begin{cases} F_Y'(y) = 1 & 0<y<1 \\ 0 & \text{otherwise} \end{cases}$ which implies $Y \sim U(0,1)$

8.3 Exponential Distribution

pf: $\boxed{f(x) = \begin{cases} \lambda e^{-\lambda x} & x>0 \\ 0 & \text{otherwise.} \end{cases}}$

setup: In a Poi process for events in time let X be the length of time we wait for the first event occurrence.

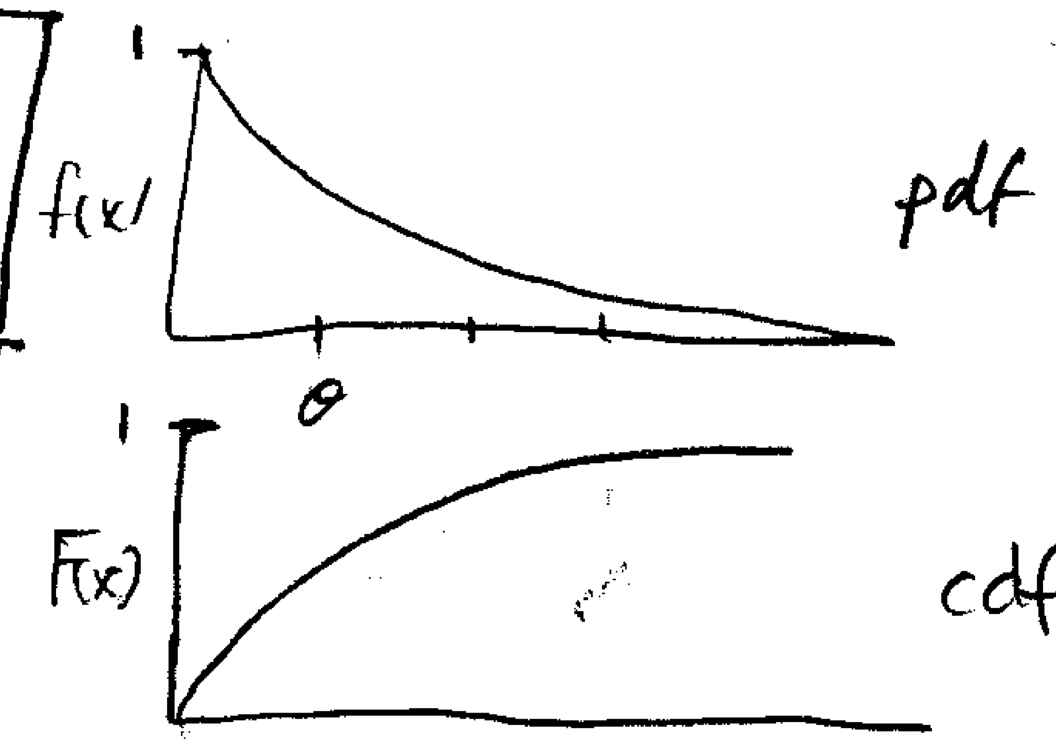
illustration: the length of time X we wait with Geiger counter until the emission of a radioactive particle is recorded follows an Exp dis.

pdf & cdf

$\boxed{\text{use } \theta = \frac{1}{\lambda}}$

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0 \\ 0 & x \le 0 \end{cases}$$



defn The Gamma Function

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

is called the gamma function of α, where $\alpha > 0$

properties:

1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$
2) $\Gamma(\alpha) = (\alpha - 1)!$ if α is a positive integer.
3) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

mean & var:

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^\infty x \frac{1}{\theta} e^{-x/\theta} dx = \theta\,\Gamma(2) = \theta(1!) = \theta$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^\infty x^2 \frac{1}{\theta} e^{-x/\theta} dx = \theta^2 \Gamma(3) = \theta^2(2!) = 2\theta^2$$

$$Var(x) = \sigma^2 = E(x^2) - \mu^2 = 2\theta^2 - \theta^2 = \theta^2$$

$\therefore$ $\boxed{\text{for } X \sim Exp(\theta) \quad E(X) = \theta \quad Var(X) = \theta^2}$

ex Suppose buses arrive at a bus stop according to a Poi process with an average of 5 buses per hour. ($\lambda = 5/\text{hour}$ $\therefore \theta = 1/5$ hour)

Find the probability:

a) you have to wait longer than 15 minutes for a bus

b) you have to wait more than 15 minutes longer, having already waited for 6 min

soln a) $P(X>15) = 1 - P(X \le 15) = 1 - F(15) = 1 - (1 - e^{-15/12}) = e^{-1.25} = 0.2865$

b) let X = total waiting time

$$P(X>21 \mid X>6) = \frac{P(X>21 \text{ and } X>6)}{P(X>6)}$$

$$= \frac{P(X>21)}{P(X>6)} = \frac{1-(1-e^{-21/12})}{1-(1-e^{-6/12})}$$

$$= e^{-1.25} = 0.2685$$

memoryless property of Exp dis:

For a ~~g~~ Poi process, given that you have waited b units of time for the next event, the prob you wait an additional c units of time does not depend on b but only depend on c.

8.5 Normal Distribution

setup: a r.v. X has a Normal dis if it has pdf of the form

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad x \in \mathbb{R}}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are parameters.

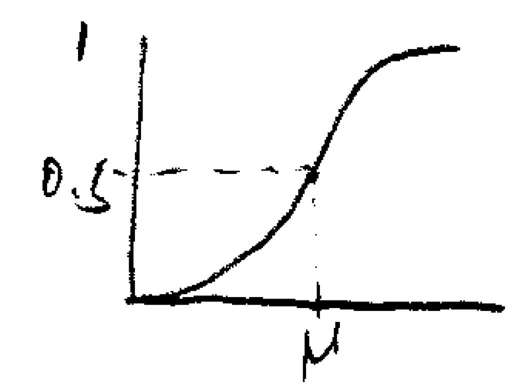
It turns out that $E(X) = \mu$ and $var(X) = \sigma^2$

We write $X \sim N(\mu, \sigma^2)$

illustration: 1) heights or weights of males (or of females) in large populations tends to follow a Normal dis

2) the logarithms of stock prices are often assumed to have Normal dis.

cdf: $\boxed{F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) dy}$



theorem let $X \sim N(\mu, \sigma^2)$ and define $Z = (X-\mu)/\sigma$, then $Z \sim N(0,1)$ and $P(X \le x) = P(Z \le \frac{x-\mu}{\sigma})$

ex find the following prob: where $Z \sim N(0,1)$

(a) $P(Z \le 2.11)$

(b) $P(Z \le 3.40)$

(c) $P(Z \ge 1.06)$

(d) $P(Z < -1.06)$

(e) $P(-1.06 < Z < 2.11)$

soln
(a) $P(Z \le 2.11) = 0.98257$

(b) $P(Z \le 3.40) = 0.99966$

(c) $P(Z > 1.06) = 1 - P(Z \le 1.06) = 1 - 0.85543 = 0.14457$

(d) $P(Z < -1.06) = P(Z > 1.06) = 1 - P(Z < 1.06) = 0.14457$

(e) $P(-1.06 < Z < 2.11) = P(Z < 2.11) - P(Z < -1.06) = 0.83800$

Finding $N(\mu, \sigma^2)$ prob: use thm which implies if $X \sim N(\mu, \sigma^2)$ then

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right)$$

where $Z \sim N(0,1)$

Gaussian Distribution the normal distribution is also known as the Gaussian distribution. The notation $X \sim G(\mu, \sigma)$ means that X has Gaussian dis with mean μ and standard deviation σ. So if $X \sim N(1, 4)$ then we could also write $X \sim G(1, 2)$

ex. The distribution of heights of adult males in Canada is well approximate by a Gaussian distribution with mean $\mu = 69.0$ inches and standard deviation $\sigma = 2.4$ inches. Find the 10^{th} and 90^{th} percentiles of the height distribution.

soln: We are being told that if X is the height of a randomly selected Canadian adult male, then $X \sim G(69.0, 2.4)$ or equivalently $X \sim N(69.0, 5.76)$

To find the 90^{th} percentile c, we use

$$P(X \le c) = P\left(\frac{X-69.0}{2.4} \le \frac{c-69.0}{2.4}\right)$$
$$= P\left(Z \le \frac{c-69.0}{2.4}\right)$$
$$= 0.9$$

From the table we see $P(Z \le 1.2816) = 0.90$ so we need

$$\frac{c-69.0}{2.4} = 1.2816$$
$$c = 72.08 \text{ inches}$$

Similarly, to find c such that $P(X \le c) = 0.10$ we find that $P(Z \le -1.2816) = 0.10$

$$\frac{c-69.0}{2.4} = -1.2816$$
$$c = 65.92$$