General Review 1

Gavin Jaeger-Freeborn

1. Probability distribution function (not the same as pdf) for random variables

pmf	probability mass function	$f(x)$	$P(X = x)$ discrete
pdf	probability density function	$f(x)$	$P(X = x)$ continuous
cdf	cumulative distribution function	$F(x)$	$P(X \leq x)$ or $\int_{-\infty}^{x} f(x)dx = F(x)$

2. Variance and standard deviation

$$V(X) = \sigma^2{}_X = E((X - \mu_X)^2) = E(X^2) - \mu^2{}_X$$

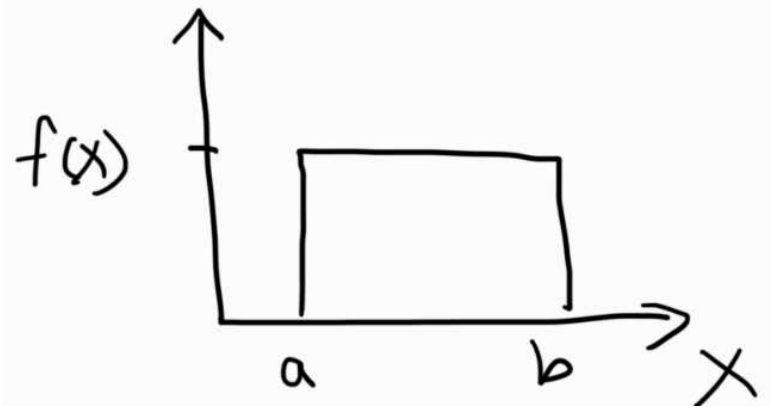
3. Uniform Distribution

Uniform distribution is a rectangle where a is the minimum and b is the maximum

$$X \sim U(a, b)$$

pmf (probability mass function)	$f(x) = \frac{1}{b-a}$
mean	$\mu = \frac{a+b}{2}$
standard deviation	$\sigma = \frac{b-a}{\sqrt{12}}$

diagram of uniform distribution



4. Poisson Distribution

If arrivals occur at random in time (or space) at the average rate of δ per unit time (or space), and X = total number of arrivals that occur in a time (or space) window of size t, then the distribution of X is: $Poisson(\lambda = \delta t)$

If $X \sim Poisson(\lambda)$, then

pmf (probability mass function)	$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$
mean	$\mu = \lambda$
standard deviation	$\sigma = \sqrt{\lambda}$

NOTE: for Poisson $\mu = \lambda$ and $\sigma^2 = \lambda$

NOTE: to use the table the probability must be of this form $P(X \leq x)$

5. Binomial Distribution

if X = total number of successes out of n independent trials where P(success) = p on every trial, then the distrobuion of X is $Binomial(n, p)$

If $X \sim Binomial(n, p)$, then

pmf (probability mass function)	$f(x) = \binom{n}{x} p^x q^{n-x}$
mean	$\mu = np$
standard deviation	$\sigma = \sqrt{npq}$

NOTE: $\binom{n}{r} = \frac{n!}{r!(n-r)!} = nCr$

NOTE: q is just the probability of failure aka $q = 1 - p$

NOTE: to use the table the probability must be of this form $P(X \leq x)$

6. Normal Distribution

Every linear combination of independent normally distributed rev's is normally distributed

If $X \sim N(\mu, \sigma)$, then the distribution of

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

is: Standard Normal

NOTE: $P(Z > z) = P(Z < -z)$ for using the table we must have $P(Z < z)$ from

7. Statistics and Distribution

let $X = X_1 + X_2 + \cdots + X_n$ be random variables. Some common statistics include:

sample mean	$\bar{X} = \frac{X_1 + X_2 \cdots + X_n}{n}$
sample sum	$T = X_1 + X_2 + \cdots + X_n$
sample variance	$S^2 = \frac{\sum(X_1 - \bar{X})^2}{n-1}$
sample median	$\tilde{X} = median(X_1, X_2, \ldots, X_n)$

NOTE: the value is $\bar{x}, t, x^2, \tilde{x}$ while $\bar{X}, T, S^2, \tilde{X}$ is a random variable.

NOTE: we want X_i~Some Distribution and have she same μ, σ^2, σ

8. Set 21 The Importance of Normal Distribution

if $X_1, X_2, \cdots, X_n$ are **independent identical** (meaning they have the same distribution) **random variables**

Then:

1. The sample mean , $\bar{X}$, has mean μ and standard deviation $\sigma/\sqrt{n}$
2. The sample sum , T, has mean $n\mu$ and standard deviation $\sigma\sqrt{n}$

   All linear combinations of independent normal random variables are normally distributed.

   If $X_1, X_2, \cdots, X_n$ are all iid (independent identical), and normally distributed then:

1. $\bar{X}$, has normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$

   $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

2. T, has normal distribution with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$

   $T \sim N(n\mu, \sigma\sqrt{n})$

   NOTE: This is for any sample size

Example

Suppose it is known that the levels of fluid in soda bottles is normally distributed, with a mean of 355 mL, and standard deviation of 2 mL. Let X_1, X_2, X_3, X_4 denote liquid content of four randomly selected bottles.

Find the probability that the average liquid content will be less than 356 mL.

$$\mu = 355ml$$

$$\sigma^2 = 2ml$$

$$n = 4$$

find

$$\bar{X} \sim N(355, 2/\sqrt{2})$$

$$P(\bar{X} < 356ml)$$

$$= P(Z < \frac{X - \mu}{\sigma})$$

$$= P(Z < \frac{356 - 355}{1})$$

$$= P(Z < 1)$$

FROM TABLE

$$0.8413$$

9. Central Limit Theorem

Let $X1, X2, \ldots, Xn$ be iid random variables, each with mean μ and standard deviation σ. Provided that $n \geq 30$(rule of thumb).

1. $\bar{X}$, has *approximately* normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

$$\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$$

2. T, has *approximately* normal distribution with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$

$$T \approx N(n\mu, \sigma\sqrt{n})$$

WE MUST KNOW WHAT μ, σ, AND n ARE

NOTE: the larger the sample size the closer $\bar{X}$ and T will be to a normal distribution.

Example

The number of bacteria per mL sample of water has a Poisson distribution, with an average of 50 bacteria per sample. Suppose that 100 samples are tested. What is the probability that the average number of bacteria per sample is at least 52?

$$\lambda = 50 \text{ per sample}$$
$$n = 100$$

Using Central Limit Theorem since $n > 30$

NOTE: when using $\bar{X}$ you dont want to use Poisson

For Poisson Distribution recall that $\mu = \lambda = 50$ and $\sigma^2 = \lambda = 50 \therefore \sigma = \sqrt{\lambda}$

$$\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$$
$$\bar{X} \approx N(50, \frac{\sqrt{50}}{\sqrt{100}})$$
$$\bar{X} \approx N(50, 0.7071)$$
$$P(\bar{X} \geq 53) \approx P(Z > \frac{52-50}{0.7071})$$
$$\approx P(Z > \frac{52-50}{0.7071})$$
$$\approx P(Z > 2.828)$$
$$\approx P(Z > 2.83)$$

Since standard normal is symmetric about 0

$$\approx P(Z < -2.83) = 0.0023$$

FROM TABLE

$$\approx 0.0023$$

Example

In a particular lake, the amount of pollutant in a 1 L sample is has a mean of 6 mg with a standard deviation of 1 mg. Suppose we take 50 randomly selected samples, each of 1 L of lake water. What is the probability that the total amount of pollutant will be between 295 mg and 305 mg?

$$n = 50$$

$$\mu = 6mg$$

$$\sigma = 1mg$$

$$P(295 < T < 305)$$

$$P(T < 305) - P(T < 295)$$

Applying CLT we get

$$T \approx N(n\mu, \sigma\sqrt{n})$$

$$T \approx N(50 \cdot 6, 1\sqrt{50})$$

$$T \approx N(300, \sqrt{50})$$

$$\approx P(Z < \frac{305 - 300}{\sqrt{50}}) - P(Z < \frac{295 - 300}{\sqrt{50}})$$

$$\approx P(Z < 0.7071) - P(Z < -0.7071)$$

$$\approx 0.7611 - 0.2389$$

$$0.5222$$

in R

```
> pnorm ( 305, 300, sqrt ( 50 ) ) - pnorm( 295 , 300, sqrt (50 ) )
[1] 0.5204999
```

Example

Pheasants in a particular region were found to have an appreciable mercury contamination. The mercury level in parts per million for these birds is normally distributed with mean 0.25 and standard deviation 0.08.

If I select 4 pheasants at random, what is the probability that the mean mercury level will be greater than 0.3 ppm?

$$n = 4$$

Since $n = 4 < 30$ we cannot use CLT

$$\mu = 0.25$$

$$\sigma = 0.08$$

$$X_1, \ldots, X_4 \sim N(0.25, 0.08)$$

$$P(\bar{X} > 0.3)$$

This is nolonger an approximation since we are using a normal distrobuion

$$\bar{X} \sim N(0.25, \frac{0.088}{\sqrt{4}} = 0.04)$$

$$P(\bar{X} > 0.3)$$

$$P(Z > \frac{0.3 - 0.25}{0.04}) = P(Z > 1.25)$$

$$= P(Z < -1.25)$$

$$= 0.1056$$

in R

```
> pnorm( -1.25 )
[1] 0.1056498
```

Example

Suppose again that we select 4 pheasants at random. What is the probability that all of the pheasants will have a mercury level which is less than 0.2?

y = # of pheasants having mercury levels < 0.2 ppm

$\therefore$ success = (having mercury levels < 0.2 ppm)

$$P = P(X < 0.2)$$

$$= P(X < \frac{0.2 - 0.25}{0.08})$$

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

$$= P(Z < -0.635)$$

$$= 0.2643$$

Now apply this to a binomial distribution

$$y \sim Bin(4, 0.2643)$$

$$P(y = 4) = \binom{4}{4}(0.2643)^4(1 - 0.2643)^0$$

$$= 0.2643^4$$

$$= 0.0049$$