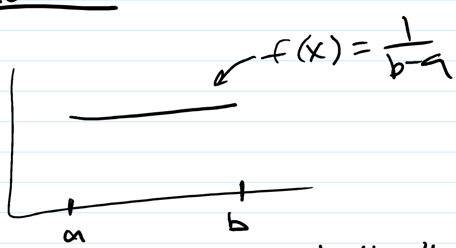
Day 14 Notes

Thursday, August 26, 2021 9:04 AM

Probability distributions

Uniform



$L \times H = (b-a) \cdot \frac{1}{b-a} = 1$

Bernoulli

Binary outcome (0/1)

Parameter: $Pr(X=1) = p$

$E(X) = p$

$Var(X) = p(1-p)$ ★

I.I.D. = independent, identically distributed

Binomial

Bernoulli: 1 coin flip

Binomial: # of heads out of K coin flips

$\mu = np$ for $Bin(n,p)$

$\sigma^2 = np(1-p)$

$Bin(n,p) = \sum_{i=1}^{n} Ber(p)$

$E[Bin] = E[\sum Ber]$

$= \sum E[Ber]$

$= \sum p$

$= \Sigma p$

$= np$

$Var[Bin] = Var[\Sigma Ber]$

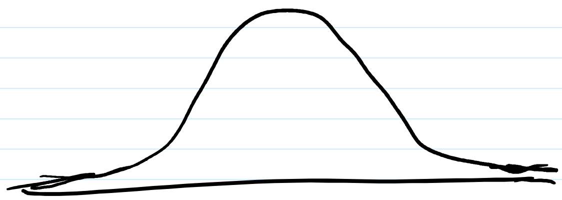
$= \Sigma Var(Ber)$ [all covariances are 0 by IID]

$= \Sigma p(1-p)$

$= np(1-p)$

Normal distribution

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



$y = \beta_0 + \underbrace{\beta_1}_{} X + \varepsilon$

estimate w/ $\hat{\beta}_1 = (X'X)^{-1}X'y$

$\hat{\beta}_1$ is a statistic w/ a (unknown) probability distribution

CLT allows us to assume

$\hat{\beta}_1$ has an approximately normal distrib. and we can separately estimate its mean and variance $\rightarrow$ and μ and σ^2 completely characterize a normal dist.

Standard Normal: $N(0,1)$ "ϕ"

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

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

Then $Y = \frac{X-\mu}{\sigma}$, $Y \sim N(0,1)$

"Z-distribution"

$$\Sigma N \sim N(\Sigma \mu, \Sigma \sigma^2)$$

→ this is unusual among probability distributions

☆ Normal dist. has very thin tails
→ outliers very rare

Exponential

Survival analysis: constant rate of failure ("Poisson process")

The interval between failures is exponentially distributed.

Sometimes λ is replaced w/ $\lambda^* = \frac{1}{\lambda}$

$\text{Exp}(\lambda) = \text{Gamma}(1, \lambda)$

~~Bivariate~~ and Multivariate Normal

Multivariate Normal is not just the joint distribution of N i.i.d. Normal variables. It has additional properties.

→ In particular, covariances can be specified.

Sampling distributions

All of the earlier distributions are applied to many diff. kinds of RV's.

A statistic is a function of several RV's (often different random draws in a sample).

Examples:

Sample mean: $\frac{\sum x_i}{n} = \bar{x}$

Sample variance: $\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

Note: $\bar{x}$ and $\hat{\sigma}^2$ are statistics.

$n-1$

Note: $\bar{x}$ and $\hat{\sigma}^2$ are <u>statistics</u>, and are themselves random variables w/ distributions.

compare to μ and σ^2 which are <u>certain</u> (if unknown) parameters, NOT r.v.'s.

<u>Ex</u> $\hat{\beta}_1$ in least squares

Law of Large Numbers

→ a property of the <u>sample mean</u>

Law of large numbers. Suppose that $X_1, \dots, X_n$ form a random sample from a distribution for which the mean is μ and for which the variance exists. Let $\overline{X}_n$ denote the sample mean. Then

$$\overline{X}_n \xrightarrow{p} \mu. \quad \textbf{(4.8.4)}$$

Proof. Let the variance of each X_i be σ^2. It then follows from the Chebyshev inequality that for every number $\varepsilon > 0$,

=>

$$\Pr(|\overline{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

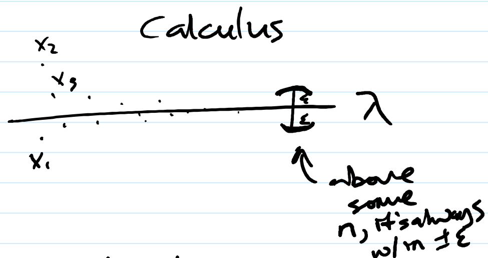
Hence,

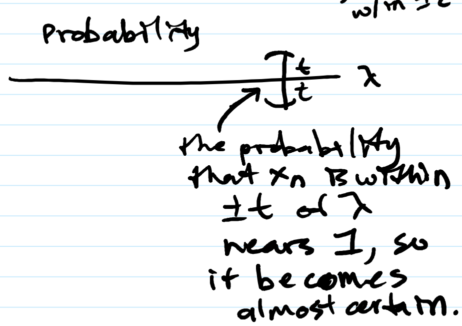
$$\lim_{n\to\infty} \Pr(|\overline{X}_n - \mu| < \varepsilon) = 1,$$

which means that $\overline{X}_n \xrightarrow{p} \mu$. ∎

convergence in probability:

> consider a sequence of random variables $\{X_n\}$. $\{X_n\}$ is said the converge in probability to λ if $\forall t > 0$,
>
> $$\lim_{n\to\infty} \Pr(|x_n - \lambda| < t) = 1$$





For each $t > 0$, for each $\varepsilon > 0$, $\exists n^*$ s.t. whenever $n > n^*$,

$$\left| \Pr(|x_n - \lambda| < t) - 1 \right| < \varepsilon$$

We say $\text{plim}\, x_n = \lambda$

or $x_n \xrightarrow{p} \lambda$

or "the probability limit of x_n is λ"

Example: Let x_n be the sample mean of a random sample. So $x_1 = y_1$

$$X_2 = \frac{y_1+y_2}{2}$$

$$X_3 = \frac{y_1+y_2+y_3}{3}$$

$$\vdots$$

Markov Inequality. Suppose that X is a random variable such that $\Pr(X \geq 0) = 1$. Then for every given number $t > 0$,

$$\Pr(X \geq t) \leq \frac{E(X)}{t}. \qquad (4.8.1)$$

Proof. For convenience, we shall assume that X has a discrete distribution for which the p.f. is f. The proof for a continuous distribution or a more general type of distribution is similar. For a discrete distribution,

$$E(X) = \sum_x xf(x) = \sum_{x<t} xf(x) + \sum_{x\geq t} xf(x).$$

Since X can have only nonnegative values, all the terms in the summations are nonnegative. Therefore,

$$E(X) \geq \sum_{x\geq t} xf(x) \geq \sum_{x \geq t} tf(x) = t\Pr(X \geq t). \qquad ■$$

Chebyshev Inequality. Let X be a random variable for which $\mathrm{Var}(X)$ exists. Then for every number $t > 0$,

$$\Pr(|X - E(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}. \qquad (4.8.2)$$

Proof. Let $Y = [X - E(X)]^2$. Then $\Pr(Y \geq 0) = 1$ and $E(Y) = \mathrm{Var}(X)$. By applying the Markov inequality to Y, we obtain the following result:

$$\Pr(|X - E(X)| \geq t) = \Pr(Y \geq t^2) \leq \frac{\mathrm{Var}(X)}{t^2}. \qquad ■$$

LLN says: $\bar{X}_n \xrightarrow{P} \mu$

$\bar{X}_n$: sample mean, a statistic

μ: population mean, a parameter

Common sampling distributions

Chi-Square or χ^2

Let $Z_1, \dots, Z_n$ be iid standard normal. Then

$$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$$

$\mu = n$

$\sigma^2 = 2n$

Let $X_1, \dots, X_k$ be independent χ^2 variables w/ parameters $n_1, n_2, \dots, n_k$, then:

$$\sum_{i=1}^{k} X_i \sim \chi^2_{(n_1+n_2+\cdots+n_k)}$$

We use χ^2 variables mainly in 3 situations:

1) sampling distribution for the sample variance of a sample from a Normal distribution

2) Null distribution for test statistics about categorical variables

2) [illegible]
statistics about categorical variables

3) Test statistics of asymptotic distributions

Is $\mu = 0$ or not?

Null hypothesis: $\mu = 5$

Alternative hyp: $\mu \neq 5$

Let's estimate μ with $\bar{X}$.

If null hypothesis were true, $\bar{x}$ would have certain distribution (centered around 5).
That's the <u>null</u> distribution.

Student's t

Let $Z \sim N(0,1)$ and let $Y \sim \chi^2_n$ then

$$\frac{Z}{\sqrt{Y/n}} \sim t(n)$$

Intuition: $\sqrt{Y/n}$

$$= \frac{z_1^2 + z_2 + \cdots + z_n^2}{n} \rightarrow \text{this looks like a variance sample}$$

$$= \sqrt{Y/n} \rightarrow \text{sample std. dev.}$$

Z: $\frac{X - \mu}{\sigma}$ if X is Normal then this is a standard normal

---

t arises as the distribution of the sample mean of a Normal random variable.

Consider an i.i.d. sample from <u>any</u> distribution with mean μ and

Consider an i.i.d. sample from <u>any</u> distribution with mean μ and variance σ^2.

Then $Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$

$$= \frac{1}{n^2}\Big(\sum_i \underbrace{var(X_i)}_{\text{each} = \sigma^2}\Big)$$

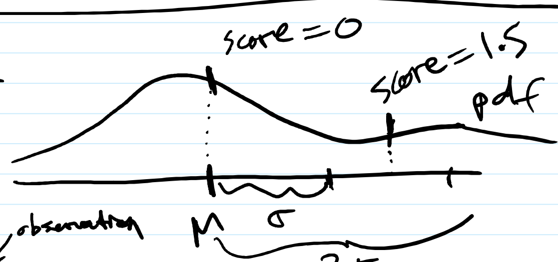
$$= \frac{1}{n^2} \cdot n\sigma^2$$

$$\boxed{Var(\bar{X}) = \frac{\sigma^2}{n}}$$

$$\boxed{StdDev(\bar{X}) = \frac{\sigma}{\sqrt{n}}}$$

Now consider a sample from a Normal distribution. We want to know the distribution of the sample mean around the actual mean.

<u>Aside</u>



$\frac{X_i - \mu}{\sigma}$ → standard score.

It tells how far X_i is from the mean in terms of "standard deviation units"

t gives the distribution of the standard scores of the sample mean: $\frac{\bar{X} - \mu}{\sigma}$

Sample mean $\bar{X}$ (known)
True mean μ (unknown)
True variance σ^2 (unknown)
Sample variance $\hat{\sigma}^2$ (known)

$E(\bar{X})$

$E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n} \cdot \sum E(X_i)$

True variance σ^2 (unknown)
Sample variance $\hat{\sigma}^2$ (known)

Standard score: $\left\{ \dfrac{\bar{X} - \mu}{\hat{\sigma}(\bar{X})} \right.$

unknown $\hat{\sigma}$

$$E(\bar{X})$$
$$E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{1}{n} \cdot \sum E(x_i)$$
$$= \frac{1}{n} \cdot n\mu$$
$$= \mu$$

$$\hat{\sigma}(\bar{X}) = \frac{\hat{\sigma}(x)}{\sqrt{n}} = \sqrt{\frac{\hat{\sigma}^2(x)}{n}}$$

Score: $\dfrac{\bar{X} - \mu}{\sqrt{\dfrac{\hat{\sigma}^2(x)}{n}}}$

This has a t-distribution w/ $(n-1)$ degrees of freedom.

t-distribs give the exact distrib. of the standard scores of a sample mean when the variables are Normal.

Linear regression: If you assume $\varepsilon_i \sim N(0, \sigma^2)$ then $\dfrac{\hat{\beta}_1 - \beta_1}{SD(\hat{\beta}_1)} \sim t_{n-k}$

If $X \sim t_n$ then $X^2 \sim F(1, n)$

F-distribution

F-distribution

Let $X \sim \chi^2_n$

and $Y \sim \chi^2_m$

be independent.

Then $\dfrac{X/n}{Y/m} \sim F(n,m)$

this arises as a ratio of sums of squares, esp. as a ratio of variances (eg. btwn to alternative models that you want to compare). These lead to test statistics w/ F-distributions.

EX compare two models

$Y_i = \beta_0 + \varepsilon_i$ model 1

vs $Y_i = \beta_0 + \underline{\beta_1} X_i + \varepsilon_i$ model 2

$$\frac{\frac{\sum \varepsilon_i'}{N}}{\frac{\sum \varepsilon_i^2}{N}} \sim F_{N,N}$$ (if ε_i are Normal or approx. normal)

Tomorrow: t-dist

CLT

asymptotic distribs.