Math 4501 - Probability and Statistics II

5.3 - Several random variables (review)

Mathematical expectation: multiple random variables

Definition

Let $X_1, \ldots, X_n$ be n random variables with joint pmf/pdf

$$f(x_1, x_2, \ldots, x_n) \quad , \quad (x_1, x_2, \ldots, x_n) \in S \ ,$$

and let $u(x_1, x_2, \ldots, x_n)$ be a real-valued function of $(x_1, x_2, \ldots, x_n)$ on S.

The mathematical expectation (or expected value) of $u(X_1, X_2, \ldots, X_n)$ is given by

$$E\left[u(X_1, X_2, \ldots, X_n)\right] = \begin{cases} \displaystyle\sum_{(x_1, \ldots, x_n) \in S} u(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) \\ \\ \displaystyle\int_S u(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n) \, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \end{cases}$$

provided the summations/integrals are absolutely convergent.

Expectation of independent random variables

Theorem (IMPORTANT!)

Let $X_1, X_2, \ldots, X_n$ be independent random variables, and let $u_1, u_2, \ldots, u_n$ be continuous real-valued functions of a single real variable.

The following hold:

1) *The random variables $u_1(X_1), u_2(X_2), \ldots, u_n(X_n)$ are also independent.*

very useful

2) *If $E\left[u_i\left(X_i\right)\right]$ exists for each $i = 1, 2, \ldots, n$, then*

$$\boxed{E\left[u_1\left(X_1\right) u_2\left(X_2\right) \cdots u_n\left(X_n\right)\right]} = E\left[u_1\left(X_1\right)\right] E\left[u_2\left(X_2\right)\right] \cdots E\left[u_n\left(X_n\right)\right]$$

Note: The continuity assumption was included solely for simplicity of exposition and the result above applies to a much broader class of functions.

Why is (1) true?

X_1, X_2 independent

$u_1(X_1)$, $u_2(X_2)$ independent

Proof sketch of (1):

Recall that $X_1, X_2, \dots, X_m$ are independent (previous lecture)

if $\Longleftrightarrow$ $P(X_1 \in A_1, X_2 \in A_2, \dots, X_m \in A_m) = P(X_1 \in A_1) . P(X_2 \in A_2) \dots P(X_m \in A_m)$

for any sets $A_1, A_2, \dots, A_m \subseteq \mathbb{R}$.

Take "arbitrary" sets $A_1, A_2, \dots, A_m \subseteq \mathbb{R}$ (open intervals would do)

and note that $P(u_1(X_1) \in A_1, u_2(X_2) \in A_2, \dots, u_m(X_m) \in A_m)$

$= P(X_1 \in u_1^{-1}(A_1), X_2 \in u^{-1}(A_2), \dots, X_m \in u^{-1}(A_m))$ $\leftarrow$

by independence $\rightarrow$ $= P(X_1 \in u_1^{-1}(A_1)) \cdot P(X_2 \in u_2^{-1}(A_2)) \cdots P(X_m \in u_m^{-1}(A_m))$

$= P(u_1(X_1) \in A_1) \cdot P(u_2(X_2) \in A_2) \cdots P(u_m(X_m) \in A_m)$

$\Rightarrow u_1(X_1), \ldots, u_m(X_m)$ are independent

Sketch of proof of 2 (on the continuous case) $\rightsquigarrow$ some of steps are similar to what we've done in the last example of our 1st class

$X_1, \ldots, X_m$ independent iff

joint pdf $f_{joint}(x_1, x_2, \ldots, x_m) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_m}(x_m)$ for all $(x_1, \ldots, x_m) \in \mathbb{R}^m$

Then $E[u_1(X_1) \cdot u_2(X_2) \cdots u_m(X_m)] = \int\int_{\mathbb{R}^m} \cdots \int u_1(x_1) u_2(x_2) \cdots u_m(x_m) \cdot f_{joint}(x_1, \ldots, x_m)\, dx_1 \ldots dx_m$

$= \cdots$

$$= \int\int_{\mathbb{R}^n} \cdots \int u_1(x_1)\, u_2(x_2) \cdots u_n(x_n)\, \overbrace{f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_n(x_n)}\, dx_1 \cdots dx_n$$

Step where we use independence

$$= \underbrace{\left(\int_{\mathbb{R}} u_1(x_1)\, f_{X_1}(x_1)\, dx_1 \right)} \cdot \underbrace{\left(\int_{\mathbb{R}} u_2(x_2) \cdot f_{X_2}(x_2)\, dx_2 \right)} \cdots \underbrace{\left(\int_{\mathbb{R}} u_n(x_n) \cdot f_{X_n}(x_n)\, dx_n \right)}$$

$$= E[u_1(X_1)] \cdot E[u_2(X_2)] \cdots \cdot E[u_n(X_n)]$$

Linear combination of random variables (review + additional detail)

Theorem (VERY IMPORTANT!)

Let $X_1, X_2, \ldots, X_n$ be random variables with respective means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$.

Let Y be given by

$$Y = \sum_{i=1}^{n} a_i X_i \ ,$$

r.v. Y is a linear combination of $X_1, X_2, \ldots, X_n$

where $a_1, a_2, \ldots, a_n$ are real constants.

The mean of Y is

$$\mu_Y = \sum_{i=1}^{n} a_i \mu_i \ .$$

: mean of linear combination equals linear combination of means

If, in addition, $X_1, X_2, \ldots, X_n$ are independent, then

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 \ .$$

Proof: set $\boxed{Y = \sum_{i=1}^{m} a_i X_i}$, $a_1, \dots, a_m \in \mathbb{R}$.

1st part

$$\mu_Y = E[Y] = E\left[\sum_{i=1}^{m} a_i X_i\right] \underset{\text{linearity}}{=} \sum_{i=1}^{m} E[a_i X_i] = \sum_{i=1}^{m} a_i\, E[X_i] = \sum_{i=1}^{m} a_i \mu_i$$

linearity of expectation

2nd part

def. of variance

$$\sigma_Y^2 = \mathrm{Var}(Y) = E\left[(Y - E[Y])^2\right] = E\left[\Big(\underbrace{\sum_{i=1}^{m} a_i X_i}_{Y} - \underbrace{\sum_{i=1}^{m} a_i \mu_i}_{E[Y] \text{ by the 1st part}}\Big)^2\right]$$

$$= E\left[\Big(\sum_{i=1}^{m} a_i (X_i - \mu_i)\Big)^2\right] = E\left[\sum_{i=1}^{m} \sum_{j=1}^{m} a_i (X_i - \mu_i)\, a_j (X_j - \mu_j)\right]$$

special case.

With 3 terms:

$$\left[a_1 (X_1 - \mu_1) + a_2 (X_2 - \mu_2) + a_3 (X_3 - \mu_3) \right]^2$$

$$= \left[a_1 (X_1 - \mu_1) + a_2 (X_2 - \mu_2) + a_3 (X_3 - \mu_3) \right] \cdot \left[a_1 (X_1 - \mu_1) + a_2 (X_2 - \mu_2) + a_3 (X_3 - \mu_3) \right]$$

each term on LHS must multiply each of the 3 terms on RHS resulting in 9 terms (which we might eventually combine)

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} a_i (X_i - \mu_i)\, a_j (X_j - \mu_j)$$

Use linearity of expectation to get

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \underbrace{E\left[(X_i-\mu_i)(X_j-\mu_j)\right]}_{Cov(X_i, X_j)}$$

$$E[X_i^2]-(E[X_i])^2 = Var(X_i) = Cov(X_i, X_i)$$

Since $X_1, X_2, \ldots, X_m$ are independent, we have that for $i \neq j$

$$Cov(X_i, X_j) = E\Big[\underbrace{(X_i-\mu_i)}_{u_i(X_i)}\underbrace{(X_j-\mu_j)}_{u_j(X_j)}\Big] = \underbrace{E[X_i-\mu_i]}_{0}\cdot\underbrace{E[X_j-\mu_j]}_{0}$$

previous thm + independence

$$= (E[X_i]-\mu_i)\cdot(E[X_j]-\mu_j) = (\mu_i-\mu_i)(\mu_j-\mu_j) = 0$$

CONCLUSION: Under independence $\sigma_Y^2 = \sum_{i=1}^{m} a_i^2 \, Cov(X_i, X_i) = \sum_{i=1}^{m} a_i^2 \, Var(X_i)$

In the absence of the independence assumption on $X_1, X_2, \ldots, X_n$, the second item of the previous theorem would be:

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j) \ ,$$

general statement without independence

where $\mathrm{Cov}(X_i, X_j)$ denotes the covariance of X_i and X_j, given by

$$\mathrm{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \ ,$$

Example

Let the independent random variables X_1 and X_2 have respective means $\mu_1 = -4$ and $\mu_2 = 3$ and variances $\sigma_1^2 = 4$ and $\sigma_2^2 = 9$.

Find the mean and the variance of $Y = 3X_1 - 2X_2$.

$E[X_1] = -4$, $E[X_2] = 3$, $Var(X_1) = 4$, $Var(X_2) = 9$

$$E[Y] = E[3X_1 - 2X_2] \underset{\substack{\text{previous thm} \\ \text{(linearity of expectation)}}}{=} 3E[X_1] - 2E[X_2] = 3(-4) - 2 \cdot 3 = -18$$

$$Var(Y) = Var(3X_1 - 2X_2) \underset{\text{independence}}{=} (3)^2 Var(X_1) + (-2)^2 Var(X_2) = 9 \cdot 4 + 4 \cdot 9 = 72$$

Example

$\underbrace{\text{random sample}}$ $\longrightarrow$ i i d independent and identically distributed

Let X_1, X_2 be a random sample from a distribution with mean μ and variance σ^2.

Find the mean and the variance of $Y = X_1 - X_2$.

We know: $E[X_1] = E[X_2] = \mu$ and $Var(X_1) = Var(X_2) = \sigma^2$

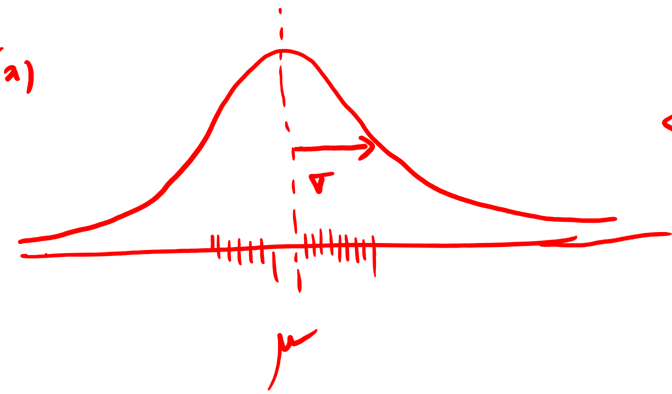
$$E[Y] = E[X_1 - X_2] \underset{\text{linearity}}{=} E[X_1] - E[X_2] = \mu - \mu = 0$$

$$Var(Y) = Var(X_1 - X_2) = (1)^2 \cdot Var(X_1) + (-1)^2 Var(X_2)$$

$$= Var(X_1) + Var(X_2) = \sigma^2 + \sigma^2 = 2\sigma^2$$

Illustration: X_1 and X_2 both have the distribution

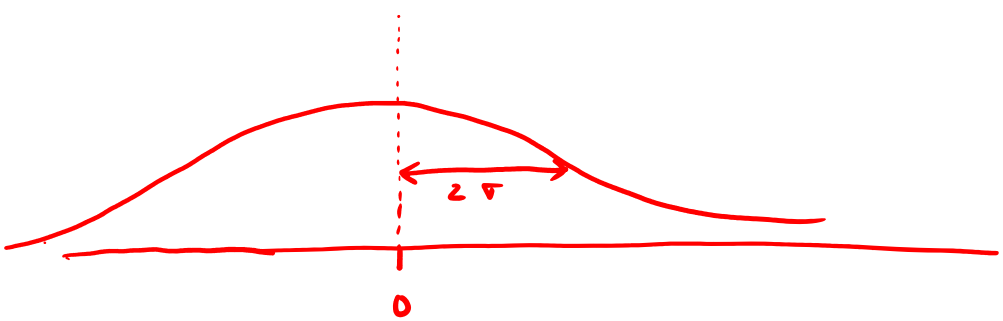
$f_X(x)$



$X_1 \to \mu+\sigma,\ \mu,\ \mu-\sigma$ AND $X_2 \to \mu+\sigma,\ \mu,\ \mu-\sigma$

$X_1 - X_2 \to 2\sigma,\ \sigma,\ 0,\ -\sigma,\ -2\sigma$

$Y = X_1 - X_2$



Statistic

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n. (i i d)

Any function of the sample observations $X_1, X_2, \ldots, X_n$ that does not have any unknown parameters is called a *statistic*.

Depends ONLY on $\rightarrow$ $x_1, x_2, \ldots, x_n$
$\rightarrow$ known quantities or parameter

function of the random sample for which we know all values of the included parameters

Sample mean and sample variance

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n.

Its *sample mean* is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and its *sample variance* is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 .$$

Example

The sample mean and the sample variance of a random sample of size n are statistics.

Mean and variance of the sample mean

no specific info on distribution besides mean and variance

Theorem

i i d

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2.

The mean and variance of the sample mean $\bar{X}$ are

$$\boxed{\mu_{\bar{X}} = E\left(\bar{X}\right) = \mu}$$

and

$$\boxed{\sigma^2_{\bar{X}} = \mathrm{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}\ .}$$

Proof :

$$\mu_{\bar{X}} = E[\bar{X}] = E\left[\frac{1}{m}\sum_{i=1}^{m} X_i\right] \underset{\uparrow \text{ linearity of } E}{=} \frac{1}{m}\sum_{i=1}^{m} \overbrace{E[X_i]}^{\mu} = \frac{1}{m}\underbrace{\sum_{i=1}^{m}\mu}_{m\mu} = \frac{1}{m}\cdot m\mu = \mu$$

$$\sigma^2_{\bar{X}} = \mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) \underset{\uparrow \text{ independence}}{=} \frac{1}{m^2}\sum_{i=1}^{m} \overbrace{\mathrm{Var}(X_i)}^{\sigma^2} =$$

$$= \frac{1}{m^2}\overbrace{\sum_{i=1}^{m}\sigma^2}^{\overbrace{\sigma^2+\sigma^2+\cdots\sigma^2}^{m \text{ time}}} = \frac{1}{m^2}\cdot(m\sigma^2) = \frac{1}{m}\sigma^2 = \frac{\sigma^2}{m}$$

Math 4501 - Probability and Statistics II

3.2 - The $\overbrace{\text{chi-square distribution}}^{\text{new content}}$

(as an element of the $\underbrace{\text{gamma distribution family}}_{\text{review}}$)

Gamma distribution (review)

A random variable X has a gamma distribution with parameters $\theta > 0$ and $\alpha > 0$ if its pdf is of the form

$$\Longrightarrow \qquad f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, \quad x > 0 .$$

special case:
if $\alpha = 1$: Gamma distribution is the exponential distr. with mean θ

Note: the waiting time W until the αth occurrence in a Poisson process with parameter λ has a gamma distribution with parameters α and

$$\theta = \frac{1}{\lambda} .$$

Gamma function (review)

generalization of factorial to non-integer values

The gamma function is defined by

$$\boxed{\Gamma(t) = \int_0^{\infty} y^{t-1} \mathrm{e}^{-y} \, \mathrm{d}y \,, \quad t > 0 \,.}$$

Properties:

1. $\Gamma(t) > 0$ for every $t > 0$
2. $\Gamma(1) = \int_0^{\infty} \mathrm{e}^{-y} \, \mathrm{d}y = 1$
3. If $t > 1$, integration by parts yields

$$\Gamma(t) = (t-1)\Gamma(t-1)$$

4. For every positive integer $n \in \mathbb{N}$, we have

$$\boxed{\Gamma(n) = (n-1)!}$$

Mgf for gamma distribution (review)

Suppose X has a gamma distribution with parameters $\theta > 0$ and $\alpha > 0$.

The mgf of X is given by

$$M(t) = \frac{1}{(1 - \theta t)^{\alpha}}, \qquad t < \frac{1}{\theta}$$

The mean and variance of X are

- $\mu = E[X] = \alpha\theta$
- $\sigma^2 = \mathrm{Var}(X) = \alpha\theta^2$

Chi-square distribution

Let X have a gamma distribution with parameters $\theta = 2$ and $\alpha = r/2$, where r is a positive integer.

The pdf of X is

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, \quad x > 0$$

and we say that X has a chi-square distribution with r degrees of freedom.

gamma with
$\theta = 2$
$\alpha = \frac{r}{2}$, $r \in \mathbb{N}$
d.o.f

Notation and terminology: We say that X is $\chi^2(r)$ and denote it as $\boxed{X \sim \chi^2(r)}$.

X follows a χ^2 distribution with r d.o.f.

Notes:

- $\chi^2(r)$ is a gamma distribution with $\theta = 2$ and $\alpha = r/2$, where $r \in \mathbb{N}$.
- $\chi^2(2)$ is an exponential distribution with $\theta = 2$.

if $r = 2 \Rightarrow \alpha = \frac{2}{2} = 1 \Rightarrow \chi^2(2)$ is $Exp(\theta = 2)$

Moment generating function for chi-square distribution

Suppose X as a chi-square distribution with r degrees of freedom.

The mgf of X is given by

$$M(t) = \frac{1}{(1-2t)^{r/2}}, \qquad t < \frac{1}{2}$$

The mean and variance of X are

- $\mu = E[X] = \frac{r}{2}2 = r$
- $\sigma^2 = \mathrm{Var}(X) = \frac{r}{2}2^2 = 2r$

Percentiles for chi-square distribution

Suppose X is $\chi^2(r)$, and let $\alpha \in (0, 1)$ (usually $\alpha < 0.5$).

The $100(1-\alpha)$th percentile is the number $\chi^2_\alpha(r)$ such that

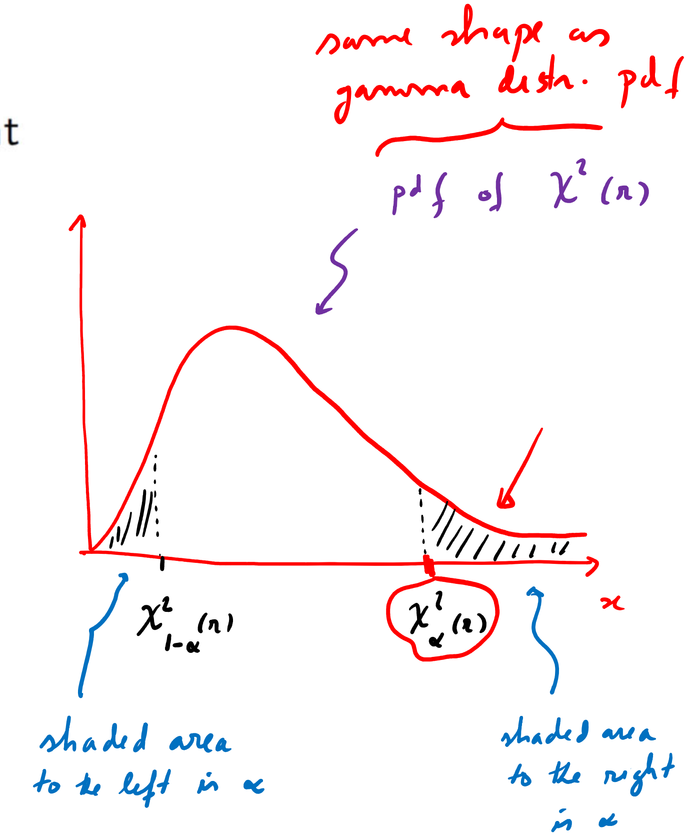
$$P\left[X \geq \chi^2_\alpha(r)\right] = \alpha \ ,$$

that is, the probability to the right of $\chi^2_\alpha(r)$ is α.

The 100α percentile is the number $\chi^2_{1-\alpha}(r)$ such that

$$P\left[X \leq \chi^2_{1-\alpha}(r)\right] = \alpha \ ,$$

that is, the probability to the right of $\chi^2_{1-\alpha}(r)$ is $1-\alpha$.



Example

Let X have a chi-square distribution with $r = 5$ degrees of freedom. Determine:

a) $P(1.145 \leq X \leq 12.83)$

b) $P(X > 15.09)$

Let $X \sim \chi^2(5)$ and denote by F the cdf of X:

pdf of $\chi^2(5)$

a) $P(1.145 \leq X \leq 12.83) = \cancel{\int_{1.145}^{12.83} f(x)\,dx}$ ← usually a hard integral to evaluate

$= F(\underline{12.83}) - F(\underline{1.145})$

NOTE: The cdf of $\chi^2(r)$ is tabled!

using table!

$= 0.975 - 0.05$

$= 0.925$ //