

Section 21

Expected Values

Po-Ning Chen, Professor

Institute of Communications Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30010, R.O.C.

Expected value as integral

21-1

Definition (expected value) The expected value of a random variable X is the integral of X with respect to its cdf. I.e.,

$$E[X] = E[X^+] - E[X^-] = \int_0^{\infty} x dF_X(x) - \int_{-\infty}^0 (-x) dF_X(x) = \int_{-\infty}^{\infty} x dF_X(x).$$

Properties of expected value

1. $E[X]$ exists if, and only if, at least one of $E[X^+]$ and $E[X^-]$ is finite.
2. X is integrable, if $E[|X|] < \infty$.
3. For \mathcal{B}/\mathcal{B} -measurable (or simply \mathcal{B} -measurable) function g ,

$$\begin{aligned} E[g(X)] &= E[g(X)^+] - E[g(X)^-] \\ &= \int_{\{x \in \mathfrak{R}: g(x) \geq 0\}} g(x) dF_X(x) - \int_{\{x \in \mathfrak{R}: g(x) < 0\}} [-g(x)] dF_X(x) \\ &= \int_{-\infty}^{\infty} g(x) dF_X(x). \end{aligned}$$

Absolute moments

21-2

Definition (absolute moments) The k th absolute moment of X is:

$$E [|X|^k] = \int_{-\infty}^{\infty} |x|^k dF_X(x).$$

Properties of absolute moments

1. The absolute moment always exists.
2. If k th absolute moment is finite, then j th moment is finite for any $j \leq k$.

Proof: It can be easily proved by $|x|^j \leq 1 + |x|^k$ for $j \leq k$. □

Definition (moments) The k th moment of X is:

$$\begin{aligned} E[X^k] &= \int_{-\infty}^{\infty} x^k dF_X(x) \\ &= \int_{-\infty}^{\infty} \max\{x^k, 0\} dF_X(x) - \int_{-\infty}^{\infty} (-\min\{x^k, 0\}) dF_X(x) \\ &= E[(X^k)^+] - E[(X^k)^-]. \end{aligned}$$

Properties of moments

1. $E[X^k]$ exists if, and only if, at least one of $E[(X^k)^+]$ and $E[(X^k)^-]$ is finite.
2. X^k is integrable, if $E[|X^k|] < \infty$.
3. For \mathcal{B}/\mathcal{B} -measurable (or simply \mathcal{B} -measurable) function g ,

$$\begin{aligned} E[g(X^k)] &= E[g(X^k)^+] - E[g(X^k)^-] \\ &= \int_{\{x \in \mathfrak{R}: g(x^k) \geq 0\}} g(x^k) dF_X(x) - \int_{\{x \in \mathfrak{R}: g(x^k) < 0\}} [-g(x^k)] dF_X(x) \\ &= \int_{-\infty}^{\infty} g(x^k) dF_X(x). \end{aligned}$$

Moments

21-4

Example 21.1 (moments of standard normal) The pdf of a standard normal distribution is equal to:

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Integration by parts shows that:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}x^k e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}(k-1)x^{k-2} e^{-x^2/2} dx,$$

or equivalently,

$$E[X^k] = (k-1)E[X^{k-2}].$$

As a result,

$$E[X^k] = \begin{cases} 0, & \text{for } k \text{ odd;} \\ 1 \times 3 \times 5 \times \cdots \times (k-1), & \text{for } k \text{ even.} \end{cases}$$

Computation of mean

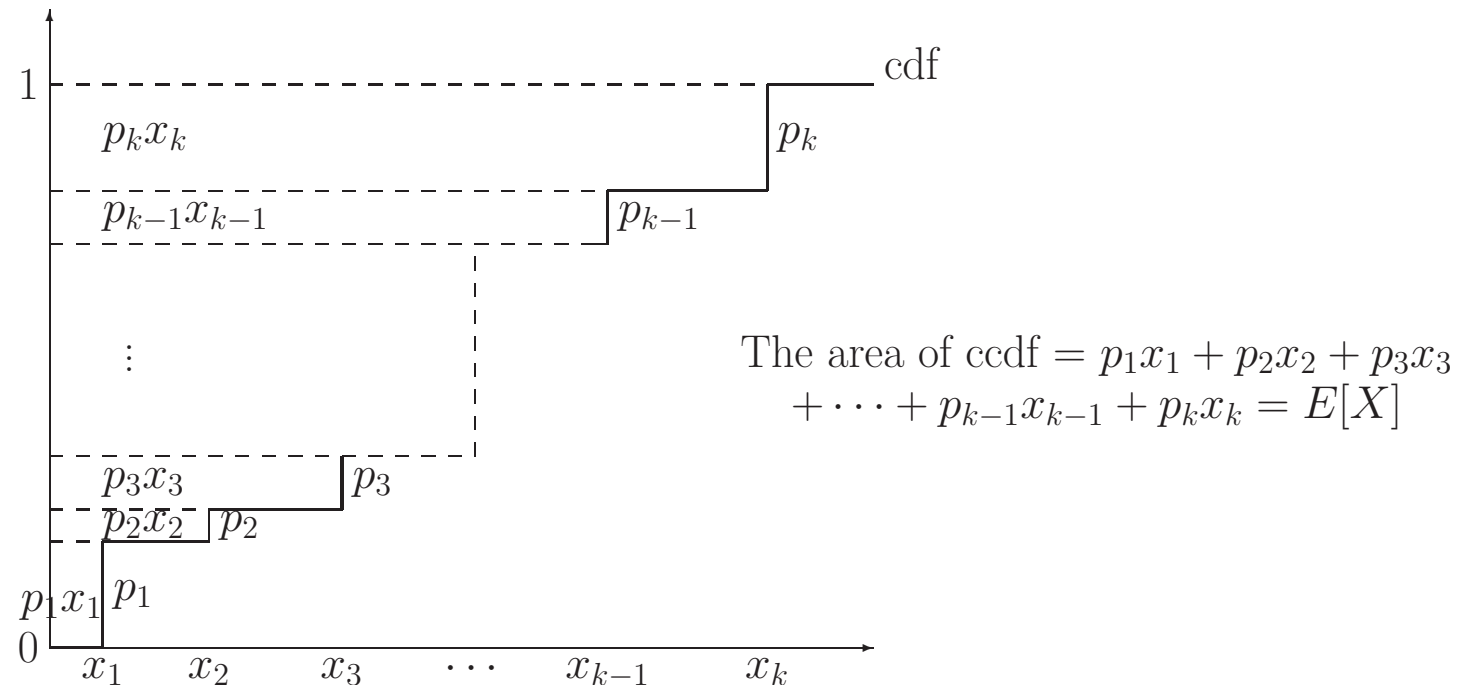
21-5

Theorem For non-negative random variable X ,

$$E[X] = \int_0^{\infty} \Pr[X > t] dt = \int_0^{\infty} \Pr[X \geq t] dt.$$

In other words, the area of cdf (complementary cdf) is the mean. (The theorem is valid even if $E[X] = \infty$.) (This is an extension of the law of large numbers when empirical distribution is concerned)!

Geometric interpretation Suppose $\Pr[X = x_i] = p_i$ for $1 \leq i \leq k$.



Note that it is always true that $\int_0^\infty x dF_X(x) = \int_{0^+}^\infty x dF_X(x)$ but it is possible $\int_\alpha^\infty x dF_X(x) \neq \int_{\alpha^+}^\infty x dF_X(x)$!

Computation of mean

21-6

Theorem For $\alpha \geq 0$,

$$\int_{\alpha^+}^\infty x dF_X(x) = \alpha \Pr[X > \alpha] + \int_\alpha^\infty \Pr[X > t] dt.$$

Proof: Let $Y = X \times I_{[X > \alpha]}$, where $I_{[X > \alpha]} = 1$ if $X > \alpha$, and zero, otherwise. Hence, $Y = 0$ for $X \leq \alpha$, and $Y = X$ for $X > \alpha$. Consequently,

$$\begin{aligned} \int_{\alpha^+}^\infty x dF_X(x) &= E[Y] = \int_0^\infty \Pr[Y > t] dt \\ &= \int_0^\alpha \Pr[Y > t] dt + \int_\alpha^\infty \Pr[Y > t] dt \\ &= \int_0^\alpha \Pr[X > \alpha] dt + \int_\alpha^\infty \Pr[X > t] dt \\ &= \alpha \Pr[X > \alpha] + \int_\alpha^\infty \Pr[X > t] dt. \end{aligned}$$

□

- The empirical approximation of $\Pr[X > t]$ (or $\Pr[X \leq t]$) is more easily obtained than $dF(x)$. With the above result, $E[X]$ (or $E[Y]$) can be established directly from $\Pr[X > t]$.

Inequalities regarding moments

21-7

Inequalities regarding moments

Markov's inequality

21-8

Lemma (Markov's inequality) For any $k > 0$ (and implicitly $\alpha > 0$),

$$\Pr[|X| \geq \alpha] \leq \frac{1}{\alpha^k} E[|X|^k].$$

Proof: The below inequality is valid for any $x \in \mathfrak{R}$:

$$|x|^k \geq \alpha^k \cdot \mathbf{1}\{|x|^k \geq \alpha^k\} \quad (21.1)$$

Hence,

$$\underline{E[|X|^k]} = \int_{-\infty}^{\infty} |x|^k dF_X(x) \geq \alpha^k \cdot \int_{-\infty}^{\infty} \mathbf{1}\{|x|^k \geq \alpha^k\} dF_X(x) = \alpha^k \underline{\Pr[|X| \geq \alpha]}.$$

Equality holds if, and only if, equality in (21.1) is true with probability 1. I.e.,

$$\Pr[|X|^k = \alpha^k \cdot \mathbf{1}\{|X|^k \geq \alpha^k\}] = 1,$$

or equivalently, $\Pr[|X| = 0 \text{ or } \alpha] = 1$. □

Chebyshev-Bienaymé inequality

21-9

Lemma (Chebyshev-Bienaymé inequality) For $\alpha > 0$,

$$\Pr[|X - E[X]| \geq \alpha] \leq \frac{1}{\alpha^2} \text{Var}[X].$$

Proof: By Markov's inequality with $k = 2$, we have:

$$\Pr[|X - E[X]| \geq \alpha] \leq \frac{1}{\alpha^2} E[|X - E[X]|^2].$$

Equality holds if, and only if,

$$\Pr[|X - E[X]| = 0] + \Pr[|X - E[X]| = \alpha] = 1,$$

which implies that

$$\Pr[X = E[X] + \alpha] = \Pr[X = E[X] - \alpha] = p$$

and

$$\Pr[X = E[X]] = 1 - 2p$$

for $\alpha > 0$.

□

Jensen's inequality

21-10

Definition (convexity) A function $\varphi(x)$ is said to be *convex* over an interval (a, b) if for every $x_1, x_2 \in (a, b)$ and $0 \leq \lambda \leq 1$,

$$\varphi(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda\varphi(x_1) + (1 - \lambda)\varphi(x_2).$$

Furthermore, a function φ is said to be *strictly convex* if equality holds only when $\lambda = 0$ or $\lambda = 1$. (Can we replace (a, b) by a real set \mathcal{X} ?)

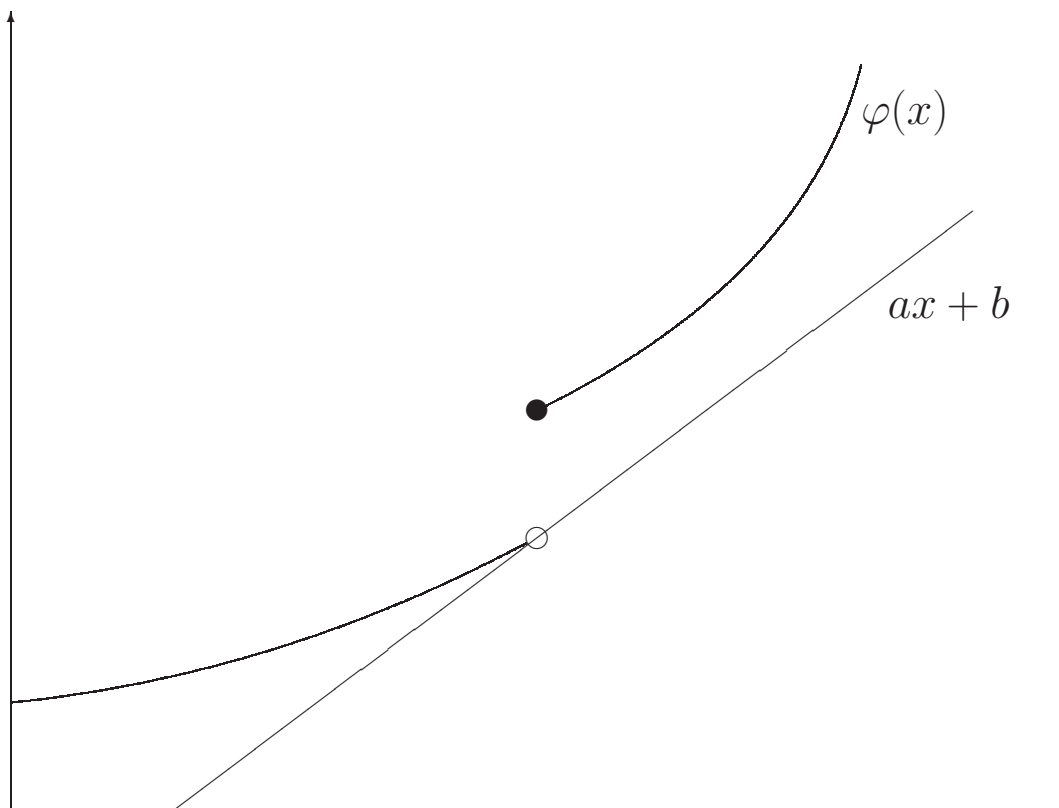
Definition (support line) A line $y = ax + b$ is said to be a support line of function $\varphi(x)$ if among all lines of the same slope a , it is the largest one satisfying $ax + b \leq \varphi(x)$ for every x .

- A support line $ax + b$ may not necessarily intersect with $\varphi(\cdot)$. In other words, it is possible that no x_0 satisfies $ax_0 + b = \varphi(x_0)$.
- However, the existence of intersection between function $\varphi(\cdot)$ and its support line is guaranteed, if $\varphi(\cdot)$ is convex.

Jensen's inequality

21-11

An example that no intersection exists for a function and its support line.



Jensen's inequality

21-12

Lemma (Jensen's inequality) Suppose that function $\varphi(\cdot)$ is convex on the domain \mathcal{X} of X . (Implicitly, $E[X] \in \mathcal{X}$.) Then

$$\varphi(E[X]) \leq E[\varphi(X)].$$

Proof: Let $ax + b$ be a support line through the point $(E[X], \varphi(E[X]))$.

Thus, over the domain \mathcal{X} of $\varphi(x)$,

$$ax + b \leq \varphi(x).$$

If equality holds in \mathcal{X} in this step, then equality remains true for the subsequent steps.

By taking the expectation value of both sides, we obtain

$$a \cdot E[X] + b \leq E[\varphi(X)],$$

but we know that $a \cdot E[X] + b = \varphi(E[X])$. Consequently,

$$\varphi(E[X]) \leq E[\varphi(X)].$$

Equality hold if, and only if, either there exist a and b such that $aE[X] + b = \varphi(E[X])$ and

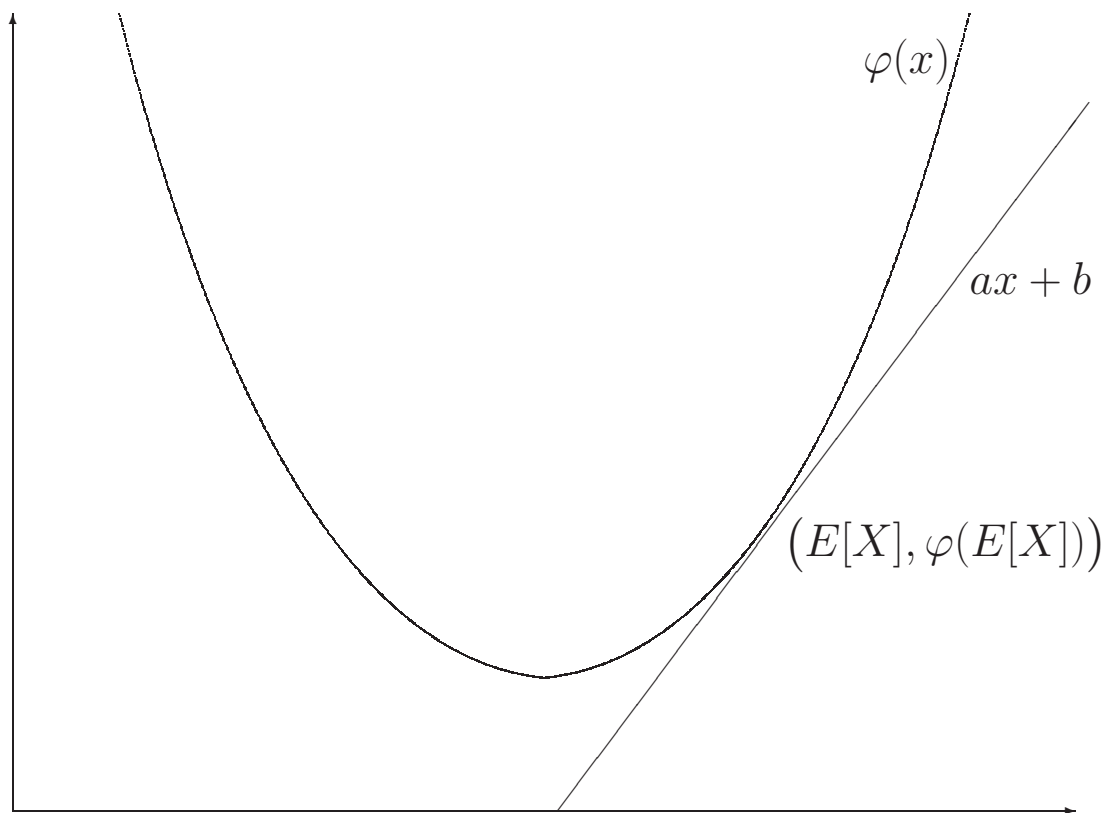
$$\Pr(\{x \in \mathcal{X} : ax + b = \varphi(x)\}) = 1.$$

□

Jensen's inequality

21-13

The support line $y = ax + b$ of the convex function $\varphi(x)$.



Hölder's inequality

21-14

Lemma (Hölder's inequality) For $p > 1, q > 1$ and $1/p + 1/q = 1$,

$$E[|XY|] \leq E^{1/p}[|X|^p]E^{1/q}[|Y|^q].$$

Proof: Since the inequality is trivially valid, if $E^{1/p}[|X|^p]E^{1/q}[|Y|^q] = 0$. Without loss of generality, assume $E^{1/p}[|X|^p]E^{1/q}[|Y|^q] > 0$.

- $\exp\{x\}$ is a convex function in x . Hence, by Jensen's inequality,

$$\exp\left\{\frac{1}{p}s + \frac{1}{q}t\right\} \leq \frac{1}{p}\exp\{s\} + \frac{1}{q}\exp\{t\}.$$

Since e^x is strictly convex, equality holds iff $s = t$.

- Let $a = \exp\{s/p\}$ and $b = \exp\{t/q\}$. Then the above inequality becomes:

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q,$$

Equality holds iff $a^p = b^q$.

whose validity is not restricted to positive a and b but to non-negative a and b .

- By letting $a = |x|/E^{1/p}[|X|^p]$ and $b = |y|/E^{1/q}[|Y|^q]$, we obtain:

$$\frac{|xy|}{E^{1/p}[|X|^p]E^{1/q}[|Y|^q]} \leq \frac{1}{p}\frac{|x|^p}{E[|X|^p]} + \frac{1}{q}\frac{|y|^q}{E[|Y|^q]}.$$

Equality holds if, and only if,
 $\Pr\left[\frac{|X|^p}{E[|X|^p]} = \frac{|Y|^q}{E[|Y|^q]}\right] = 1$.

Hölder's inequality

21-15

Taking the expectation values of both sides yields:

$$\frac{E[|XY|]}{E^{1/p}[|X|^p]E^{1/q}[|Y|^q]} \leq \frac{1}{p} \frac{E[|X|^p]}{E[|X|^p]} + \frac{1}{q} \frac{E[|Y|^q]}{E[|Y|^q]} = \frac{1}{p} + \frac{1}{q} = 1. \quad \square$$

Lemma (Hölder's inequality) For $p > 1, q > 1$ and $1/p + 1/q = 1$,

$$E[|XY|] \leq E^{1/p}[|X|^p]E^{1/q}[|Y|^q].$$

Equality holds if, and only if,

$$\Pr \left[\frac{|X|^p}{E[|X|^p]} = \frac{|Y|^q}{E[|Y|^q]} \right] = 1 \text{ or } \Pr[X = 0] = 1 \text{ or } \Pr[Y = 0] = 1.$$

Example. $p = q = 2$ and

	$Y = 0$	$Y = 1$
$X = 0$	p_{00}	p_{01}
$X = 1$	p_{10}	p_{11}

$$\begin{aligned}
 E[|XY|] &= p_{11} \\
 &= p_{11}^{1/2} p_{11}^{1/2} \\
 &\leq (p_{10} + p_{11})^{1/2} (p_{01} + p_{11})^{1/2} \\
 &= E^{1/2}[|X|^2] E^{1/2}[|Y|^2]
 \end{aligned}$$

with equality holding if, and only if, $p_{10} = p_{01} = 0$. □

Cauchy-Schwartz's inequality

21-16

Suppose $E[|X|] > 0$ and $E[|Y|] > 0$.

Lemma (Hölder's inequality) For $p > 1, q > 1$ and $1/p + 1/q = 1$,

$$E[|XY|] \leq E^{1/p}[|X|^p]E^{1/q}[|Y|^q].$$

Equality holds if, and only if, there exists a such that $\Pr[|X|^p = a|Y|^q] = 1$.

Lemma (Cauchy-Schwartz's inequality)

$$E[|XY|] \leq E^{1/2}[X^2]E^{1/2}[Y^2].$$

Equality holds if, and only if, there exists a such that $\Pr[X^2 = aY^2] = 1$.

Proof: A special case of Hölder's inequality with $p = q = 2$.

Equality holds if, and only if,

$$\Pr[X^2 = aY^2] = 1$$

for some a .

□

Lyapounov's inequality

21-17

Lemma (Lyapounov's inequality) For $0 < \alpha < \beta$,

$$E^{1/\alpha}[|Z|^\alpha] \leq E^{1/\beta}[|Z|^\beta].$$

Equality holds if, and only if, $\Pr[|Z| = a] = 1$ for some a .

Proof: Letting $X = |Z|^\alpha$, $Y = 1$, $p = \beta/\alpha$ and $q = \beta/(\beta - \alpha)$ in Hölder's inequality yields:

$$E[|Z|^\alpha] \leq E^{\alpha/\beta} \left[(|Z|^\alpha)^{\beta/\alpha} \right] E^{(\beta-\alpha)/\beta} \left[1^{\beta/(\beta-\alpha)} \right] = E^{\alpha/\beta}[|Z|^\beta].$$

Equality holds if, and only if,

$$\Pr \left[(|Z|^\alpha)^{\beta/\alpha} = a \right] = \Pr \left[|Z|^\beta = a \right] = \Pr \left[|Z| = a^{1/\beta} \right] = 1$$

for some a . □

- Notably, in the statement of the lemma, β is strictly larger than α .
- It is certain that if $\alpha = \beta$, the inequality automatically becomes an equality.

Joint integrals

21-18

Lemma For $\mathcal{B}^k/\mathcal{B}$ -measurable (or simply \mathcal{B}^k -measurable) function g and k -dimensional random vector \mathbf{X} ,

$$E[g(\mathbf{X})] = \int_{\mathfrak{R}^k} g(x^k) dF_{\mathbf{X}}(x^k),$$

if one of $E[(g(\mathbf{X}))^+]$ and $E[(g(\mathbf{X}))^-]$ is finite.

Definition (covariance) The *covariance* of two random vectors \mathbf{X} and \mathbf{Y} is:

$$\begin{aligned} \text{Cov}[\mathbf{X}, \mathbf{Y}] &= E [(\mathbf{X} - E[\mathbf{X}])(\mathbf{Y} - E[\mathbf{Y}])^T] \\ &= E \left[\begin{bmatrix} X_1 - E[X_1] \\ X_2 - E[X_2] \\ \vdots \\ X_k - E[X_k] \end{bmatrix} \begin{bmatrix} Y_1 - E[Y_1] & Y_2 - E[Y_2] & \cdots & Y_\ell - E[Y_\ell] \end{bmatrix} \right] \\ &= \begin{bmatrix} (X_1 - E[X_1])(Y_1 - E[Y_1]) & \cdots & (X_1 - E[X_1])(Y_\ell - E[Y_\ell]) \\ \vdots & \cdots & \vdots \\ (X_k - E[X_k])(Y_1 - E[Y_1]) & \cdots & (X_k - E[X_k])(Y_\ell - E[Y_\ell]) \end{bmatrix}_{k \times \ell} \end{aligned}$$

where “ T ” represents vector transpose operation, if one of $E[((X_i - E[X_i])(Y_j - E[Y_j]))^+]$ and $E[((X_i - E[X_i])(Y_j - E[Y_j]))^-]$ is finite for every i, j .

Definition (uncorrelated) \mathbf{X} and \mathbf{Y} is uncorrelated, if

$$\text{Cov}[\mathbf{X}, \mathbf{Y}] = \mathbf{0}_{k \times \ell}.$$

Definition (independence) \mathbf{X} and \mathbf{Y} is independent, if

$$\begin{aligned} & \Pr [(X_1 \leq x_1 \wedge \cdots \wedge X_k \leq x_k) \wedge (Y_1 \leq y_1 \wedge \cdots \wedge Y_\ell \leq y_\ell)] \\ &= \Pr [X_1 \leq x_1 \wedge \cdots \wedge X_k \leq x_k] \Pr [Y_1 \leq y_1 \wedge \cdots \wedge Y_\ell \leq y_\ell]. \end{aligned}$$

Lemma (integrability of product) For independent X_1, X_2, \dots, X_k , if each of X_i is integrable, so is $X_1 X_2 \cdots X_k$, and

$$E[X_1 X_2 \cdots X_k] = E[X_1] E[X_2] \cdots E[X_k].$$

Lemma (sum of variance for pair-wise independent samples) If X_1, X_2, \dots, X_k are pair-wise independent and integrable,

$$\text{Var}[X_1 + X_2 + \cdots + X_k] = \text{Var}[X_1] + \text{Var}[X_2] + \cdots + \text{Var}[X_k].$$

- Notably, *pair-wise independence* does not imply *complete independence*.

Joint integrals

21-20

Example (only pair-wise independence) Toss a fair coin twice, and assume independence. Define

$$\begin{aligned} X &= \begin{cases} 1, & \text{if head appears on the first toss;} \\ 0, & \text{otherwise,} \end{cases} \\ Y &= \begin{cases} 1, & \text{if head appears on the second toss;} \\ 0, & \text{otherwise,} \end{cases} \\ Z &= \begin{cases} 1, & \text{if exactly one head and one tail appear on the two tosses;} \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

Then $\Pr[X = 1 \wedge Y = 1 \wedge Z = 1] = 0$;

but $\Pr[X = 1] \Pr[Y = 1] \Pr[Z = 1] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$.

So X, Y and Z are not independent.

Obviously, $X \perp\!\!\!\perp Y$. In addition,

$$\left. \begin{aligned} \Pr[X = 1|Z = 1] &= \frac{\Pr[X = 1 \wedge Z = 1]}{\Pr[Z = 1]} = \frac{1/4}{1/2} = \frac{1}{2} = \Pr[X = 1] \\ \Pr[X = 1|Z = 0] &= \frac{\Pr[X = 1 \wedge Z = 0]}{\Pr[Z = 0]} = \frac{1/4}{1/2} = \frac{1}{2} = \Pr[X = 1] \end{aligned} \right\} \Rightarrow X \perp\!\!\!\perp Z.$$

One can similarly show (or by symmetry) that $Y \perp\!\!\!\perp Z$.

Joint integrals

21-21

Example (con't) By

$$\left. \begin{array}{l} \text{head head} \Rightarrow X + Y + Z = 2 \\ \text{head tail} \Rightarrow X + Y + Z = 2 \\ \text{tail head} \Rightarrow X + Y + Z = 2 \\ \text{tail tail} \Rightarrow X + Y + Z = 0 \end{array} \right\} \Rightarrow \begin{array}{l} \text{Var}[X + Y + Z] \\ = \frac{3}{4}(2 - 3/2)^2 + \frac{1}{4}(0 - 3/2)^2 = \frac{3}{4} \end{array}$$

This is equal to:

$$\text{Var}[X] + \text{Var}[Y] + \text{Var}[Z] = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

This result matches that “The variance of sum equals the sum of variances” holds for pair-wise independent random variables.

Moment generating function

21-22

Definition (moment generating function) The moment generating function of X is defined as:

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} dF_X(x),$$

for all t for which this is finite.

- If $M_X(t)$ is defined (i.e., finite) throughout an interval $(-t_0, t_0)$, where $t_0 > 0$. then

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k].$$

In other words, $M_X(t)$ has a Taylor expansion about $t = 0$ with positive radius of convergence **if it is defined in some $(-t_0, t_0)$** .

- In case that $M_X(t)$ has a Taylor expansion about $t = 0$ with positive radius of convergence, the moment of X can be computed by the derivatives of $M_X(t)$ through:

$$M^{(k)}(0) = E[X^k].$$

- If $M_{X_i}(t)$ is defined throughout an interval $(-t_0, t_0)$ for each i , and X_1, X_2, \dots, X_n are independent, then the moment generating function of $X_1 + \dots + X_n$ is also defined on $(-t_0, t_0)$, and is equal to $\prod_{i=1}^n M_{X_i}(t)$.

Moment generating function

21-23

Example (random variable whose moment generating function is defined only at zero) The pdf of a Cauchy distribution is

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

For $t \neq 0$, the integral

$$\begin{aligned} \int_{-\infty}^{\infty} e^{tx} \frac{1}{\pi(1+x^2)} dx &= \int_0^{\infty} (e^{tx} + e^{-tx}) \frac{1}{\pi(1+x^2)} dx \\ &\geq \int_0^{\infty} e^{|t|x} \frac{1}{\pi(1+x^2)} dx \\ &\geq \int_0^{\infty} \frac{|t|x}{\pi(1+x^2)} dx \quad (\text{by } e^x \geq 1+x \geq x \text{ for } x \geq 0) \\ &\geq \int_1^{\infty} \frac{|t|x}{\pi(1+x^2)} dx \geq \int_1^{\infty} \frac{|t|x}{\pi(x^2+x^2)} dx = \frac{|t|}{2\pi} \int_1^{\infty} \frac{1}{x} dx = \infty. \end{aligned}$$

So the moment generating function is not defined for any $t \neq 0$.

The Cauchy distribution is indeed the Student's T -distribution with 1 degree of freedom.

Moment generating function

21-24

Example (Student's T -distribution with n degree of freedom)

This distribution has moments of order $\leq n - 1$ but any higher moments either do not exist or are infinity.

Some books considers infinite moments as “non-existence”. So they write “This distribution has moments of order $\leq n - 1$ but no higher moments exist.”

Let X, Y_1, Y_2, \dots, Y_n be i.i.d. with standard normal marginal distribution. Then

$$T_n = \frac{X\sqrt{n}}{\sqrt{Y_1^2 + \dots + Y_n^2}}$$

is called the Student's t -distribution (on \mathfrak{R}) with n degree of freedom.

- The numerator $X\sqrt{n}$ has a normal density with mean 0 and variance n .
- $\chi^2 = Y_1^2 + \dots + Y_n^2$ is a *chi-square distribution* with n degree of freedom, and has density $f_{1/2, n/2}(y)$, where

$$f_{\alpha, \nu}(x) = \frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x} \quad \text{on } [0, \infty)$$

is the *gamma density* (or sometimes named *Erlangian density* when ν is a positive integer) with parameters $\nu > 0$ and $\alpha > 0$, and $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ is the gamma function.

Moment generating function

21-25

(closure under convolutions for gamma density) $f_{\alpha,\mu} * f_{\alpha,\nu} = f_{\alpha,\mu+\nu}$.

Proof:

$$\begin{aligned}(f_{\alpha,\mu} * f_{\alpha,\nu})(x) &= \int_0^\infty f_{\alpha,\mu}(x-y)f_{\alpha,\nu}(y)dy \\ &= \frac{\alpha^{\mu+\nu}}{\Gamma(\mu)\Gamma(\nu)} \int_0^x (x-y)^{\mu-1}e^{-\alpha(x-y)}y^{\nu-1}e^{-\alpha y}dy \\ &= \frac{\alpha^{\mu+\nu}}{\Gamma(\mu)\Gamma(\nu)}e^{-\alpha x} \int_0^x (x-y)^{\mu-1}y^{\nu-1}dy \\ &= \frac{\alpha^{\mu+\nu}}{\Gamma(\mu)\Gamma(\nu)}e^{-\alpha x} \int_0^1 (x-xt)^{\mu-1}(xt)^{\nu-1}xdt \quad (\text{by } y = xt) \\ &= \frac{1}{\Gamma(\mu+\nu)}\alpha^{\mu+\nu}x^{\mu+\nu-1}e^{-\alpha x} \cdot \underbrace{\frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\Gamma(\nu)} \int_0^1 (1-t)^{\mu-1}t^{\nu-1}dt}_{=1} \\ &= f_{\alpha,\mu+\nu}(x),\end{aligned}$$

where “that the last term is equal to one” follows the beta integral. \square

For $\mu > 0$ and $\nu > 0$, $B(\mu, \nu) = \int_0^1 (1-t)^{\mu-1}t^{\nu-1}dt = \int_0^\infty \frac{t^{\mu-1}}{(1+t)^{\mu+\nu}}dt = \frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}$ is the so-called *beta integral*.

Moment generating function

21-26

That the pdf of chi-square distribution with 1 degree of freedom equals $f_{1/2,1/2}(y)$ can be derived as:

$$\Pr[Y_1^2 \leq y] = \Pr[-\sqrt{y} \leq Y_1 \leq \sqrt{y}] = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1,$$

where $\Phi(\cdot)$ represents the unit Gaussian cdf. So the derivative of the above equation is $y^{-1/2}\phi(y^{1/2}) = \frac{1}{\Gamma(1/2)}(1/2)^{1/2}y^{(1/2)-1}e^{-(1/2)y}$ for $y \geq 0$.

Then by closure under convolution for gamma densities, the distribution of chi-square distribution with n degree of freedom can be obtained.

In statistical mechanics, $Y_1^2 + Y_2^2 + Y_3^2$ appears as the square of the speed of particles. So the density of particle speed (on $[0, \infty)$) is equal to $v(x) = 2x f_{1/2,3/2}(x^2)$, which is called *Maxwell density*.

By letting $\bar{Y} = \sqrt{Y_1^2 + \cdots + Y_n^2}$,

$$\begin{aligned} \Pr[T_n \leq t] &= \Pr[(X\sqrt{n})/\bar{Y} \leq t] = \int_0^\infty \Pr[X \leq yt/\sqrt{n}]dF_{\bar{Y}}(y) \\ &= \int_0^\infty \Phi(yt/\sqrt{n})(2y f_{1/2,n/2}(y^2))dy. \end{aligned}$$

Moment generating function

21-27

So the density of T_n is:

$$\begin{aligned} f_{T_n}(t) &= \int_0^\infty \left(\frac{y}{\sqrt{n}} \phi(yt/\sqrt{n}) \right) (2y f_{1/2, n/2}(y^2)) dy \\ &= \int_0^\infty \left(\frac{y}{\sqrt{2\pi n}} e^{-y^2 t^2 / (2n)} \right) \left(\frac{1}{2^{n/2-1} \Gamma(n/2)} y^{n-1} e^{-y^2/2} \right) dy \\ &= \frac{1}{2^{(n-1)/2} \Gamma(n/2) \sqrt{\pi n}} \int_0^\infty y^n e^{-y^2(1+t^2/n)/2} dy \\ &= \frac{1}{2^{(n-1)/2} \Gamma(n/2) \sqrt{\pi n}} \int_0^\infty \frac{2^{n/2} s^{n/2}}{(1+t^2/n)^{n/2}} e^{-s} \left(\frac{2^{-1/2} s^{-1/2}}{(1+t^2/n)^{1/2}} \right) ds \quad \left(\text{by } y = \frac{\sqrt{2s}}{(1+t^2/n)^{1/2}} \right) \\ &= \frac{1}{\Gamma(n/2) \sqrt{\pi n} (1+t^2/n)^{(n+1)/2}} \int_0^\infty s^{(n+1)/2-1} e^{-s} ds \\ &= \frac{C_n}{(1+t^2/n)^{(n+1)/2}}, \text{ where } C_n = \frac{\Gamma((n+1)/2)}{\Gamma(n/2) \sqrt{\pi n}}. \end{aligned}$$

Moment generating function

21-28

For $t \neq 0$, the integral

$$\begin{aligned} \int_{-\infty}^{\infty} e^{tx} \frac{C_n}{(1+x^2/n)^{(n+1)/2}} dx &\geq C_n \int_0^{\infty} e^{|t|x} \frac{1}{(1+x^2/n)^{(n+1)/2}} dx \\ &\geq \frac{C_n}{n!} \int_0^{\infty} \frac{|t|^n x^n}{(1+x^2/n)^{(n+1)/2}} dx \\ &\quad (\text{by } e^x \geq 1 + x + \frac{x^2}{2} + \dots + \frac{x^n}{n!} \geq \frac{x^n}{n!} \text{ for } x \geq 0) \\ &\geq \frac{C_n}{n!} \int_{\sqrt{n}}^{\infty} \frac{|t|^n x^n}{(1+x^2/n)^{(n+1)/2}} dx \\ &\geq \frac{C_n}{n!} \int_{\sqrt{n}}^{\infty} \frac{|t|^n x^n}{(x^2/n + x^2/n)^{(n+1)/2}} dx = \frac{C_n |t|^n n^{(n+1)/2}}{n! 2^{(n+1)/2}} \int_{\sqrt{n}}^{\infty} \frac{1}{x} dx = \infty. \end{aligned}$$

So the moment generating function is not defined for any $t \neq 0$.

Moment generating function

21-29

However for $r < n$,

$$\begin{aligned}
 E[|T_n|^r] &= \int_{-\infty}^{\infty} C_n \frac{|t|^r}{(1+t^2/n)^{(n+1)/2}} dt \\
 &= 2C_n \int_0^{\infty} \frac{t^r}{(1+t^2/n)^{(n+1)/2}} dt = C_n n^{(r+1)/2} \int_0^{\infty} \frac{s^{(r+1)/2-1}}{(1+s)^{(r+1)/2+(n-r)/2}} ds \\
 &= C_n n^{(r+1)/2} \frac{\Gamma((r+1)/2)\Gamma((n-r)/2)}{\Gamma((n+1)/2)} = n^{r/2} \frac{\Gamma((n-r)/2)}{\Gamma(n/2)\sqrt{\pi}} < \infty.
 \end{aligned}$$

But when $r = n$ (similarly for $r > n$),

$$E[(T_n^r)^+] = E[(T_n^n)^+] = \left\{ \begin{array}{l} \int_{-\infty}^{\infty} C_n \frac{t^n}{(1+t^2/n)^{(n+1)/2}} dt, \text{ if } n \text{ even;} \\ \int_0^{\infty} C_n \frac{t^n}{(1+t^2/n)^{(n+1)/2}} dt, \text{ if } n \text{ odd.} \end{array} \right\} = \infty$$

and

$$E[(T_n^r)^-] = E[(T_n^n)^-] = \begin{cases} 0, & \text{if } n \text{ even;} \\ \int_{-\infty}^0 C_n \frac{-t^n}{(1+t^2/n)^{(n+1)/2}} dt = \infty, & \text{if } n \text{ odd.} \end{cases}$$

This is a good example for which the moments, even if some of them exist (and are finite), cannot be obtained through the moment generating function.

Some well-known densities

21-30

gamma distribution : $f_{\alpha,\nu}(x)$ (on $[0, \infty)$) has mean ν/α and variance ν/α^2 .

Snedecor's distribution or F -distribution : It is the distribution of

$$Z = \frac{\frac{X_1^2 + \dots + X_m^2}{m}}{\frac{Y_1^2 + \dots + Y_n^2}{n}},$$

where X_i and Y_j are all independent standard normal. Its pdf with positive integer parameters m and n is:

$$\frac{m^{m/2} \Gamma((m+n)/2)}{n^{m/2} \Gamma(m/2)\Gamma(n/2)} \frac{z^{m/2-1}}{(1+mz/n)^{(m+n)/2}} \quad \text{on } z \geq 0.$$

Bilateral exponential distribution : It is the distribution of $Z = X_1 - X_2$, where X_1 and X_2 are independent and have common exponential density $\alpha e^{-\alpha x}$ on $x \geq 0$.

Its pdf is $\frac{1}{2}\alpha e^{-\alpha|x|}$ on $x \in \mathfrak{R}$.

It has zero mean and variance $2\alpha^{-2}$.

Some well-known densities

21-31

Beta distribution : Its pdf with parameters $\mu > 0$ and $\nu > 0$ is

$$\beta_{\mu,\nu}(x) = \frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)}(1 - x)^{\mu-1}x^{\nu-1} \quad \text{on } 0 < x < 1.$$

Its mean is $\nu/(\mu + \nu)$, and its variance is $\mu\nu/[(\mu + \nu)^2(\mu + \nu + 1)]$.

Arc sine distribution : Its pdf is $\beta_{1/2,1/2}(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ on $0 < x < 1$.

Its cdf is given by $\frac{2}{\pi}\sin^{-1}(\sqrt{x})$ for $0 < x < 1$.

Generalized arc sine distribution : It is a beta distribution with $\mu + \nu = 1$.

Pareto distribution : It is the distribution of $Z = X^{-1} - 1$, where X is beta distributed with parameters $\mu > 0$ and $\nu > 0$. Its density is

$$\frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)} \frac{z^{\mu-1}}{(1+z)^{\mu+\nu}} \quad \text{on } 0 < z < \infty.$$

This is often used as an incoming traffic with heavy tail as $z^{-(\nu+1)}$.

Some well-known densities

21-32

Cauchy distribution : It is the distribution of $Z = \alpha X/|Y|$, where X and Y are independent standard normal distributed.

Its pdf with parameter $\alpha > 0$ is

$$\gamma_{\alpha}(x) = \frac{1}{\pi} \frac{\alpha}{x^2 + \alpha^2} \quad \text{on } \mathfrak{R}.$$

Its cdf is $\frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x/\alpha)$.

It is also **closure under convolution**, i.e., $\gamma_s * \gamma_u = \gamma_{s+u}$.

It is interesting that Cauchy distribution is also **closure under scaling**, i.e., $a \cdot X$ has density $\gamma_{a\alpha}(\cdot)$, if X has density $\gamma_{\alpha}(\cdot)$.

Hence, we can easily obtain the density of $a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ as $\gamma_{a_1 \alpha_1 + a_2 \alpha_2 + \cdots + a_n \alpha_n}(\cdot)$, if X_i has density $\gamma_{\alpha_i}(\cdot)$.

Some well-known densities

21-33

One-sided stable distribution of index 1/2 : It is the distribution of $Z = \alpha^2 X^{-2}$, where X has a standard normal distribution.

Its pdf is $f_\alpha(z) = \frac{\alpha}{\sqrt{2\pi}} \frac{1}{\sqrt{z^3}} e^{-\frac{1}{2}\alpha^2/z}$ on $z \geq 0$.

It is also **closure under convolution**, namely, $f_\alpha * f_\beta = f_{\alpha+\beta}$.

If Z_1, Z_2, \dots, Z_n are i.i.d. with marginal density $f_\alpha(\cdot)$, then $\frac{Z_1 + \dots + Z_n}{n^2}$ also has density $f_\alpha(\cdot)$.

Some well-known densities

21-34

Weibull distribution : Its pdf and cdf are respectively given by

$$\frac{\alpha}{\beta} \left(\frac{y}{\beta} \right)^{\alpha-1} e^{-(y/\beta)^\alpha} \quad \text{and} \quad 1 - e^{-(y/\beta)^\alpha}$$

with $\alpha > 0$, $\beta > 0$ and support $(0, \infty)$.

By defining $X = 1/Y$, where Y has the above distribution, we derive the pdf and cdf of X as respectively:

$$\alpha\beta(\beta y)^{-1-\alpha} e^{-1/(\beta y)} \quad \text{and} \quad e^{-1/(\beta x)^\alpha}.$$

This is useful for ordered statistics.

For example, let $X_{(n)}$ denote the largest one among i.i.d. X_1, X_2, \dots, X_n .

Then $\Pr \left[\frac{X_{(n)}}{n} \leq x \right] = e^{-(\alpha/\pi)(1/x)}$, if X_j is Cauchy distributed with parameter α .

Or $\Pr \left[\frac{X_{(n)}}{n^2} \leq x \right] = e^{-(\alpha\sqrt{2}/\sqrt{\pi})(1/x^{1/2})}$, if X_j is a one-sided stable distribution of index $1/2$ with parameter α .

Logistic distribution : Its cdf with parameters $\alpha > 0$ and $\beta \in \Re$ is $\frac{1}{1 + e^{-\alpha x - \beta}}$ on \Re .

Distribution with nice moment generating function

21-35

Gaussian distribution :

$$M(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-m)^2/(2\sigma^2)} dx = e^{mt + \sigma^2 t^2/2},$$

which exists for all $t \in \mathfrak{R}$.

Exponential distribution :

$$M(t) = \int_0^{\infty} e^{tx} \alpha e^{-\alpha x} dx = \frac{\alpha}{\alpha - t} = \sum_{k=0}^{\infty} \left(\frac{k!}{\alpha^k} \right) \frac{t^k}{k!}$$

is defined for $t < \alpha$.

So the k th moment is $k! \alpha^{-k}$.

Poisson distribution :

$$M(t) = \sum_{r=0}^{\infty} e^{rt} e^{-\lambda} \frac{\lambda^r}{r!} = e^{\lambda(e^t - 1)},$$

which exists for all $t \in \mathfrak{R}$.