Preliminary material STK4011-f17: Statistical Inference Theory

Random variables

Definition A random variable is a function from a sample space into the real numbers

$$X : S \to \mathbb{R}$$

S is a probability space so

i) $P(A) \geq 0, A \in \mathcal{B}$

ii) $P(S) = 1$

iii) If $A_1, A_2, \ldots \in \mathcal{B}, A_i \cap A_j = \emptyset, i \neq j, then$
$P(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} P(A_i)$

Example Assume S finite. Let

$$P(A) = \frac{\#\{s : s \in A\}}{\#\{s : s \in S\}} = \frac{\text{"}favorable\text{"}}{\text{"}possible\text{"}}$$

$P_X(X = x) = P(\{s : X(s) = x\})$ is the **distribution of X** defined in terms of S. ■

Example $S = \{0, 1\}^{10}$, so $s = (s_1, \ldots, s_{10})$ where $s_i = 0$ or $s_i = 1$, $\#S = 1024$.
The probabiliy is defined by 1024 non-negative numbers with sum equal to 1.
Let $X(s) = \Sigma_{i=1}^{10} s_i$. Then $P_X(X = x)$ is the sum of those numbers corresponding to the $s = (s_1, \ldots, s_{10})$ where exactly x s_i are 1 and $n - x$ are 0.
In the special case where all the 1024 mumbers are equal and hence $1/1024$, $X \sim Binomial(10, 0.5)$.

In general, also for countable and infinite range of X, denoted by $\mathcal{X}$ when $A \subset \mathcal{X}$

$$P_X(X \in A) = P(\{s \in S : X(s) \in A)$$

Distributions can be described by the **cumulative distribution function , cdf.**

$$F_X(x) = P_X(X \leq x)$$

Remark that $F_X(x)$ is right continous.

If $P_X(X = x) = 0$ for all x, F_X is continous and X is a continous random variable.

If X is discrete, i.e. has finite or countable range, F_X is a step function.

For X discrete, the **probability mass function, pmf,** is given by $f_X(x) = P_X(X = x)$.

If F_X is continuous and differentiable, the **probability density function, pdf** satisfies

$$F_X(x) = \int_{-\infty}^{x} f(t)dt \text{ for all x}$$

The expectation of the random variable X

$$E(X) = \begin{cases} \Sigma_{x \in \mathcal{X}} x f_X(x) & \text{if X is discrete} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{if X has pdf } f_X \end{cases}$$

describes the center of the distribution.

The variance $Var(X) = E[X^2] - [E(X)]^2$ describes the spread of the distribution.

Notation:

$X \sim F_X$, X has cdf F_X

$X \sim f_X$, X has pmf/pdf f_X

$X \sim Y$ X and Y have the same distribution.

Transformations

X random variable implies that $Y = g(X)$ is a random variable where $g : \mathcal{X} \rightarrow \mathcal{Y}$,
$\mathcal{X} = \{x : f_X(x) > 0\}, \mathcal{Y} = \{y : Y = g(x) \text{ some x }\}$

Distribution of Y?

$$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A))$$

where $g^{-1}(A) = \{x : g(x) \in A\}$.

X discrete:
$f_Y(y) = P_Y(Y = y) = \Sigma_{x \in g^{-1}(y)} P_X(X = x) = \Sigma_{x \in g^{-1}(y)} f_X(x)$

X continous:
$F_Y(y) = P_Y(Y \leq y) = P_X(g(X) \leq y) =$
$P_X(X \in \{x : g(x) \leq y\}) = \int_{\{x : g(x) \leq y\}} f_X(x)$

Problem if g is not monotone.

Example

$X \sim U[0,1]$ $Y = \sin^2(X)$

Then

$\{y :\leq y_0\} = \{x : 0 \leq x \leq x_1\} \cup \{x : x_2 \leq x \leq x_3\} \cup \{x : x \geq x_4\}$

and

$$P_Y(Y \leq y_0) = P_X(X \leq x_1\} + P_X(x_2 \leq X \leq x_3) + P_X(X \geq x_4)$$

where

$$\sin^2(x_1) = \sin^2(x_2) = \sin^2(x_3) = \sin^2(x_4) = y_0$$

Consider the case where g is strictly monotone,

i.e. either strictly decreasing or increasing

The support of X is $\mathcal{X} = \{x : f_X(x) > 0\}$

$\mathcal{Y} = \{y : Y = g(x)$ some x $\}$

Theorem

Let $X \sim f_X$ and $Y = g(X)$ where g is strictly monotone. If g is strictly monotone (on the support) and g^{-1} is continuously differentiable on $\mathcal{Y}$ then

$$g_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\partial}{\partial y} g_Y{}^{-1}(y)| & y \in \mathcal{Y} \\ 0 & \textit{else} \end{cases}$$

Proof. For g increasing

$$\begin{aligned} G_Y(y) &= P(Y \leq y) = P_X(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)). \end{aligned}$$

Differentiating with respect to y

$$g_Y(y) = \frac{\partial}{\partial y} G(y) = f_X(g^{-1}(y)) \frac{\partial}{\partial y} g_Y^{-1}(y)$$

For g decreasing

$$\begin{aligned} G_Y(y) &= P(Y \leq y) = P_X(g(X) \leq y) \\ &= P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)). \end{aligned}$$

Differentiating with respect to y

$$g_Y(y) = \frac{\partial}{\partial y} G(y) = -f_X(g^{-1}(y)) \frac{\partial}{\partial y} g_Y{}^{-1}(y)$$

Example.

Theorem

Let f be a pdf and $\mu \in \mathbb{R}$ $\sigma > 0$.
Then $X \sim \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}) \Leftrightarrow \exists Z,\ Z \sim f$ and $X = \sigma Z + \mu$

Proof $\Leftarrow$: Let $g(x) = \sigma x + \mu$. Then g is monotone, $g^{-1}(x) = \frac{x-\mu}{\sigma}$ and $\frac{\partial}{\partial x} g^{-1}(x) = 1/\sigma$. If $X = g(Z)$ then $f_X(x) = f(g^{-1}(x)) \frac{\partial}{\partial x} g^{-1}(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$

$\Rightarrow$: Let $g(x) = \frac{x-\mu}{\sigma}$, $Z = g(X)$. Then $g^{-1}(z) = \sigma z + \mu$, $\frac{\partial}{\partial x} g^{-1}(x) = \sigma$, so

$$f_Z(z) = f_X(g^{-1}(z)) \frac{\partial}{\partial x} g^{-1}(z) = f(\frac{\sigma z + \mu - \mu}{\sigma}) \frac{1}{\sigma} \sigma = f(z).$$

Also $Z = \frac{X-\mu}{\sigma}$ so $X = \sigma Z + \mu$.

Example. $Y = X^2$.

$$\begin{aligned} F_Y(y) &= P_X(X^2 \le y) = P_X(-\sqrt{y} \le X \le \sqrt{y})) \\ &= P_X(X \le \sqrt{y}) - P_X(X \le -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

so

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) - (-\frac{1}{2\sqrt{y}} f_X(-\sqrt{y})) \\ &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \end{aligned}$$

This procedure can be generalized to more than two intervals.

If F_X is not strictly increasing, care is needed in the definition of F_X^{-1}.

Lemma

Let $F_X^{-1}(y) = \inf\{u : F_X(u) \geq y\}$.
Then $F_X^{-1}(F_X(x)) = x$.

Proof. By definition

$$F_X^{-1}(F_X(x)) = \inf\{u : F_X(u) \geq F_X(x)\} = x$$

where the last equality follows since F_X is right continuous.

Important transformation:

Theorem

Let X have a continuous cdf F_X and let $Y = F_X(X)$.
Then $Y \sim U[0,1]$.

Proof.
$P_Y(Y \leq y) = P_X(F_X(X) \leq y) = P_X(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y))$

since F_X^{-1} is by definition increasing. From previous Lemma

$P_X(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) = P_X(X \leq F^{-1}(y)) =$
$F_X(F_X^{-1}(y))$.

But $F_X(F_X^{-1}(y)) = y$ since F_X is continuous.

Moment generating functions.

Definition The moment generating function, mgf, of a random variable X is the function

$$M_X(t) = E[\exp(tX)]$$

provided the expectation exists in an interval $(-h, h)$.

If differentiation and integration can be interchanged

$$\begin{aligned}
\frac{\partial}{\partial t} M_X(t)|_{t=0} &= E[X\exp(tX)]|_{t=0} = E(X) \\
\frac{\partial^2}{\partial t^2} M_X(t)|_{t=0} &= E[X^2\exp(tX)]|_{t=0} = E(X^2), \\
&\text{etc.}
\end{aligned}$$

Knowledge of moments is not enough to

detererermine distribution in general.

But, if $M_X(t)$ and $M_Y(t)$ exists,

and $M_X(t) = M_Y(t), \ \forall t \in (-h, h)$ for some $h > 0$,

then $X \sim Y$.

Also, let $\{X_j\}$ be a sequence of random variables with mgf M_{X_j}

Suppose:

i) $M_{X_j}(t) \to M_X(t) \;\; \forall t \in (-h, h)$ for some $h > 0$

ii) $M_X(t)$ is a mgf for a random variable X,

then $F_{X_j}(x) \to F_X(x)$ if F_X is continuous at x

and the moments of X is given by M_X.

Common discrete distributions

Hypergeometric

Binomial

Poisson

Negative binomial

Common continuous distributions

Uniform:

$$f_X(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{else} \end{cases}$$

Gamma:

The **gamma** function is defined by the integral $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ which is finite for $\alpha > 0$.

$$f_X(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},\ 0 < x < \infty,\ \alpha > 0, \beta > 0.$$

Normal:

$$f_X(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)},\ -\infty < x < \infty.$$

Beta:

The **beta** function is defined as

$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$

$f_X(x|\alpha, \beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1,\ \alpha > 0, \beta > 0.$

Cauchy:

$f_X(x|\mu, \sigma) = \frac{1}{\sigma\pi(1+(\frac{x-\mu}{\sigma})^2)},\ -\infty < x < \infty$

Lognormal:

$\log X \sim n(\mu, \sigma^2).$

Double exponential:

$f_X(x|\mu, \sigma) = \frac{1}{2\sigma}e^{|x-\mu|/\sigma}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$

Bivariate distributions

Bivariate random variable: $\begin{pmatrix} X \\ Y \end{pmatrix} : S \to \mathbb{R}^2$

S probability space.

Discrete case: $P(X = x, Y = y) = f(x, y)$ is **the joint pmf.** $f_X(x) = P(X = x) = \Sigma_y f(x, y)$, $f_Y(y) = P(Y = y) = \Sigma_x f(x, y)$, are **the marginal pmf.**

Continuous case: $f(x, y)$ is **the joint pdf** if $P((X, Y) \in A) = \int \int_A f(x, y) dxdy$, $A \in \mathbb{R}^2$. $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$ are **the marginal pdf's.**

Example Assume $f(x,y) = 6xy^2,\ (x,y) \in (0,1)^2$.

Then

$$\begin{aligned} P(X+Y \geq 1) &= \int_0^1 \int_{1-y}^1 6xy^2 dxdy = \int_0^1 y^2 [\int_{1-y}^1 6xdx]dy \\ &= \int_0^1 y^2 [3(1-(1-y)^2]dy = 3\int_0^1 (2y^3 - y^4)dy \\ &= \frac{6}{4} - \frac{3}{5} = \frac{30-12}{20} = \frac{9}{10} \end{aligned}$$

Conditional distributions

Discrete case: $P(Y = y|X = x) = \frac{P(Y=y,X=x)}{P(X=x)} = \frac{f(x,y)}{f_X(x)}$ is **the discrete conditional pmf.**

Continuous case: $f(y|x) = \frac{f(x,y)}{f_X(x)}$ when $f_X(x) > 0$ is **the conditional pdf.**

Example $f(x, y) = e^{-y},\ 0 < x < y < \infty$

$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy = \int_{x}^{\infty} e^{-y}dy = e^{-x}.$

$f(y|x) = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)},\ x < y.$

Independence

If $f(x,y) = f_X(x)f_Y(y)$, X and Y are **independent**.

Theorem

It is sufficient that $f(x,y)$ factorizes, i.e. $f(x,y) = g(x)h(y)$ for some functions g and h for X and Y to be independent.

Proof From the marginal densities

$$f_X(x) = g(x)\int_{-\infty}^{\infty} h(y)dy = g(x)d \text{ so } \int_{-\infty}^{\infty} h(y)dy = d$$

$$f_Y(y) = h(y)\int_{-\infty}^{\infty} g(x)dx = h(y)c \text{ so } \int_{-\infty}^{\infty} g(x)dx = c.$$

From the simultaneous density $1 = \int_{-\infty}^{\infty} f(x,y)dxdy$

$= \int_{-\infty}^{\infty} g(x)h(y)dxdy = (\int_{-\infty}^{\infty} g(x)dx)(\int_{-\infty}^{\infty} h(y)dy) = cd.$
Hence $f(x,y) = g(x)h(y)\cdot 1 = g(x)h(y)cd = f_X(x)f_Y(y)$.

Example

$f(x,y) = \frac{1}{384}x^2y^4e^{-y-x/2}$ factorizes so we do not need to compute $f_X(x) = \frac{1}{16}x^2e^{-x/2}, i.e. X \sim \chi^2_6$ and $f_Y(y) = \frac{1}{24}y^4e^{-y}, i.e. Y \sim gamma(5,1)$ to check independence.

Theorem

If X and Y are independent, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for all functions g and h for which the expectation exist.

Proof

$$\begin{aligned}
E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)f_X(x)h(y)f_Y(y)dxdy \\
&= (\int_{-\infty}^{\infty} g(x)f_X(x)dx)(\int_{-\infty}^{\infty} h(y)f_Y(y)dy) \\
&= E[g(X)]E[h(Y)]
\end{aligned}$$

Remark that $f(x,y)$ does not have to factorize for all x, y to establish independence. It is sufficient that $\int\int_B f(x,y) = 0$ on the exceptional set B.

Example

$$(X, Y) \quad \text{have pdf} \quad f(x,y) = \begin{cases} e^{-x-y} & x, y > 0 \\ 0 & \text{else} \end{cases}$$

$$(X^*, Y^*) \quad \text{have pdf} \quad f^*(x,y) = \begin{cases} e^{-x-y} & x, y > 0, \ x \neq y \\ 0 & \text{else} \end{cases}$$

Then

$$\begin{aligned} P((X,Y) \in A) &= \int\int_A f(x,y)dxdy \\ &= \int\int_A f^*(x,y)dxdy = P((X^*,Y^*) \in A) \end{aligned}$$

so $(X, Y) \sim (X^*, Y^*)$.

The point is that the pdf is only unique up to a set with probability zero.

Bivariate transformations

Suppose (X, Y) has a known distribution and $g_1 : \mathbb{R}^2 \to \mathbb{R}$, $g_2 : \mathbb{R}^2 \to \mathbb{R}$ be two specified function. Then

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} g_1(X, Y) \\ g_2(X, Y) \end{pmatrix}$$

is another bivariate random variable whose distribution is determined by the distribution of (X, Y).

Distribution of (U, V)?

Discrete case: $f(u, v) = \Sigma_{\{(x,y):g_1(x,y)=u,g_2(x,y)=v\}} f(x, y)$.

Continuous case: Let

$\mathcal{A} = \{(x, y) : f(x, y) > 0\}$
$\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ some}(x, y) \in \mathcal{A}\}$

Consider a generalzation of the treatment of univariate transformations, i.e. the case where the transformation $(g_1, g_2) : \mathcal{A} \to \mathcal{B}$ is 1-1 and onto. Then it can be inverted so $x = h_1(u, v), y = h_2(u, v)$ for all $(u, v) \in \mathcal{B}$.

The role played by the derivative in the univariate case is now played by the *Jacobian of the transformation*, $\mathcal{J}$, defined as the determinant of the matrix of partial derivatives, i.e.

$$\mathcal{J} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \det \begin{pmatrix} \frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v} \\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v} \end{pmatrix}$$

so $\mathcal{J} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$.

Theorem

Suppose the transformation $(g_1, g_2) : \mathcal{A} \to \mathcal{B}$ *is 1-1 and onto and that the Jacobian is not identically zero. Then it can be inverted so* $x = h_1(u, v), y = h_2(u, v)$ *for all* $(u, v) \in \mathcal{B}$*. The joint pdf of* (U, V) *is*

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|\mathcal{J}|$$

where $\mathcal{J}$ *is the Jacobian of the transformation.*

If the transformation is not 1-1, the theorem can be generalized by splitting $\mathcal{A}$ into sets where the transformation is 1-1.

Example Suppose $X \sim gamma(\alpha_1, \beta)$ and $Y \sim gamma(\alpha_2, \beta)$

Define $U = X + Y, V = \frac{X}{X+Y}$, so $V = \frac{X}{U}, X = UV$ and $Y = U - UV = U(1 - V)$.

Jacobian: $\mathcal{J} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} = v(-u) - u(1 - v) = -u$.

$$\begin{aligned} f_{XY}(x, y) &= \frac{1}{\Gamma(\alpha_1)}(\frac{x}{\beta})^{\alpha_1 - 1}\frac{1}{\beta}e^{-x/\beta}\frac{1}{\Gamma(\alpha_2)}(\frac{y}{\beta})^{\alpha_2 - 1}\frac{1}{\beta}e^{-y/\beta} \\ &= \frac{1}{\Gamma(\alpha_1)}\frac{1}{\Gamma(\alpha_2)}x^{\alpha_1 - 1}y^{\alpha_2 - 1}(\frac{1}{\beta})^{\alpha_1 + \alpha_2}e^{-(x+y)/\beta} \end{aligned}$$

and

$$\begin{aligned} f_{UV}(u, v) &= \frac{1}{\Gamma(\alpha_1)}\frac{1}{\Gamma(\alpha_2)}(uv)^{\alpha_1 - 1}(u(1 - v))^{\alpha_2 - 1}(\frac{1}{\beta})^{\alpha_1 + \alpha_2}e^{-u/\beta}u \\ &= \frac{1}{\Gamma(\alpha_1)}\frac{1}{\Gamma(\alpha_2)}v^{\alpha_1 - 1}(1 - v)^{\alpha_2 - 1}(\frac{u}{\beta})^{\alpha_1 + \alpha_2 - 1}\frac{1}{\beta}e^{-u/\beta} \end{aligned}$$

Hence

$$f_{UV}(u,v) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1-1}(1-v)^{\alpha_2-1} \times \frac{1}{\Gamma(\alpha_1+\alpha_2)}(\frac{u}{\beta})^{\alpha_1+\alpha_2-1}\frac{1}{\beta}e^{-u/\beta}.$$

so U and V are independent, $U \sim gamma(\alpha_1+\alpha_2)$ and $V \sim beta(\alpha_1,\alpha_2)$.

Example Suppose $X, Y \sim N(0,1)$ and independent. Let

$U = \frac{X}{Y} = g_1(X, Y)$ and $V = X = g_2(X, Y)$ so

$Y = \frac{X}{U} = \frac{V}{U} = h_1(U, V),\ X = V = h_2(U, V).$

$$\begin{aligned} f_{UV}(u, v) &= \frac{1}{2\pi} e^{-v^2(1+\frac{1}{u^2})/2} \left| \det \begin{pmatrix} 0 & 1 \\ -\frac{v}{u^2} & \frac{1}{u} \end{pmatrix} \right| \\ &= \frac{1}{2\pi} e^{-v^2(1+\frac{1}{u^2})/2} \left| \frac{v}{u^2} \right| \end{aligned}$$

Marginal distribution of U:

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-v^2(1+\frac{1}{u^2})/2} \frac{|v|}{u^2} dv \\ &= \int_{0}^{\infty} \frac{1}{2\pi} e^{-v^2(1+\frac{1}{u^2})/2} \frac{v}{u^2} dv \end{aligned}$$

Change of variable $y = v\sqrt{1+\frac{1}{u^2}}$ so $v = \frac{y}{\sqrt{1+\frac{1}{u^2}}}$,

$$\begin{aligned} f_U(u) &= \int_{0}^{\infty} \frac{1}{2\pi} e^{-y^2/2} \frac{y}{\sqrt{1+\frac{1}{u^2}}} \frac{1}{u^2} \frac{1}{\sqrt{1+\frac{1}{u^2}}} dy \\ &= \frac{1}{\pi} \frac{1}{1+u^2} \int_{0}^{\infty} y e^{-y^2/2} dy = \frac{1}{\pi} \frac{1}{1+u^2} \end{aligned}$$

since $\int_0^\infty y e^{-y^2/2} dy = -e^{-y^2/2}|_{y=0}^{y=\infty} = 1$, so $U \sim Cauchy(0,1)$.

Multivariate distributions

Multivariate random variable: $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} : S \to \mathbb{R}^n$

S probability space.

Results from bivariate case, $n = 2$, generalize

- pmf/pdf: $f(x_1, \cdots, x_n)$
- expectation: $E[g(x)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \cdots, x_n) f(x_1, \cdots, x_n) dx_1 \cdots dx_n$
- marginal pmf/pdf: obtained by summation/integration
- conditional pmf/pdf: $f(x_{k+1}, \cdots, x_n | x_1, \cdots, x_k) = \frac{f(x_1, \cdots, x_n)}{f(x_1, \cdots, x_k)}$
- mgf: $M_X(t_1, \cdots, t_n) = E[e^{\sum_{j=1}^{n} t_j X_j}]$

Multivariate transformations

Let $\mathcal{A} = \{(x_1, \ldots, x_n) : f(x_1, \ldots, x_n) > 0\}$
Suppose that the function

$$g(x) = \begin{pmatrix} g_1(x_1, \ldots, x_n) \\ \vdots \\ g_n(x_1, \ldots, x_n) \end{pmatrix} : \mathcal{A} \to \mathbb{R}^n$$

is 1-1 and onto
$\mathcal{B} = \{(u_1, \ldots, u_n) | u_j = g_j(x_1, \ldots, x_n)\}$ for some $x = (x_1, \ldots, x_n)' \in \mathcal{A}\}$.

Then there exist inverses $x_j = h_j(u_1, \ldots, u_n),\ j = 1, \ldots n$.

If the Jacobian

$$\det\begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} \end{pmatrix} = \det\begin{pmatrix} \frac{\partial h_1(u_1,\ldots,u_n)}{\partial u_1} & \cdots & \frac{\partial h_1(u_1,\ldots,u_n)}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial h_n(u_1,\ldots,u_n)}{\partial u_1} & \cdots & \frac{\partial h_n(u_1,\ldots,u_n)}{\partial u_n} \end{pmatrix}$$

is not identically 0 on $\mathcal{B}$,

$$f_U(u_1,\ldots,u_n) = f_X(h_1(u_1,\ldots,u_n),\ldots,h_n(u_1,\ldots,u_n))|\mathcal{J}|.$$

Example Suppose $X = (X_1, X_2, X_3, X_4)'$ has joint pdf

$$f_X(x_1, x_2, x_3, x_4) = \begin{cases} 24e^{-(x_1+x_2+x_3+x_4)}, & 0 < x_1 < x_2 < x_3 < x_4 < \infty \\ 0 & \text{else} \end{cases}$$

Let $U_1 = X_1, U_2 = X_2 - X_1, , U_3 = X_3 - X_2, U_4 = X_4 - X_3$.

The transformation is 1-1 and onto from
$\{(x_1, x_2, x_3, x_4) | 0 < x_1 < x_2 < x_3 < x_4 < \infty\}$ to $\mathbb{R}^{+4}$.

The inverse is
$x_1 = u_1, x_2 = u_1 + u_2, x_3 = u_1 + u_2 + u_3, x_4 = u_1 + u_2 + u_3 + u_4$
with Jacobian

$$\det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} = 1$$

so

$$f_u(u_1, u_2, u_3, u_4) = \begin{cases} 24e^{-(4u_1+3u_2+2u_3+u_4)}, & 0 < u_j < \infty, all\ j \\ 0 & \text{else} \end{cases}$$

$$= \begin{cases} 4e^{-4u_1}3e^{-u_2}2e^{-u3}e^{-u_4}, & 0 < u_j < \infty, all\ j \\ 0 & \text{else} \end{cases}$$

Also f_U factorizes so $U_1, U_2, U_3 U_4$ are independent.

Random Samples

Definition: The random variables $X_1, \ldots, X_n$ are called a **random sample** from the population $f(x)$ if $X_1, \ldots, X_n$ are mutually independent and identically distributed with pmf/pdf f(x).

Alternatively:

- $X_1, \ldots, X_n$ are i.i.d $X_j \sim f(x)$, $j = 1, \ldots, n$, or
- $X_1, \ldots, X_n$ is a sample from an infinite population.

Finite population: Sampling with replacement yields a random sample. Sampling without replacement does not yield a random sample because of dependence.

Sums of r.v. fram a random sample

After sampling there are realizations $X_1 = x_1, \dots, X_n = x_n$ which can be summarized using various measures $y = T(x_1, \dots, x_n)$, e.g. mean, median.

Regarded as a random variable $Y = T(X_1, \dots, X_n)$ is called a **statistic**. The proability distribution of Y is called the **sampling distribution**.

Two important statistics:

- Mean: $\bar{X} = \frac{1}{n}\Sigma_{j=1}^{n} X_j$
- Sample variance: $S^2 = \frac{1}{n-1}\Sigma_{j=1}^{n}(X_j - \bar{X})^2$

Then

* $E(\bar{X}) = \mu = (\int_{-\infty}^{\infty} xf(x)dx)$
* $Var(\bar{X}) = \frac{\sigma^2}{n} = (\frac{1}{n}\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx)$
* $E(S^2) = \sigma^2$

so $\bar{X}$ and S^2 are unbiased.

Sample distribution of $\bar{X}$?

Two approaches:

(i) mgf: If $X_1, \ldots, X_n$ is a random sample and $M_X(t) = E(e^{tX_j})$, $j = 1, \ldots, n$, then $M_{\bar{X}}(t) = [M_X(t/n)]^n$.

Problems are that $M_X(t)$ does not always exist and also the density can be difficult to recognize.

(ii) convolution:

Theorem

If X and Y are independent, $X \sim f_X(x)$ and $Y \sim f_Y(y)$ and $Z = X + Y$, then $f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z-u) du$.

Proof
Let $U = X$, $Z = X + Y$.

The inverse transformation is $X = U$, $Y = Z - U$. The Jacobian is

$$\mathcal{J} = \det \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

so

$$f_{UZ}(u, z) = f_X(u) f_Y(z - u)$$

and

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z - u) du.$$

Example

$U \sim Cauchy(0, \sigma)$, $V \sim Cauchy(0, \tau)$, U, V independent.

$U + V \sim Cauchy(\sigma + \tau)$ (after some calculations) If $Z_1, \ldots, Z_n$

random sample $Cauchy(0, 1)$, $\bar{Z} \sim Cauchy(0, 1)$

$\bar{X}$ in location/scale families: If $X_1, \ldots, X_n$ random sample,

$X_j \sim \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}),\ j = 1, \ldots, n.$

Then $X_j = \sigma Z_j + \mu$ and $\bar{X} = \sigma \bar{Z} + \mu$.

Hence, if $\bar{Z} \sim g(z)$,

$$\bar{X} \sim \frac{1}{\sigma} g(\frac{x-\mu}{\sigma}).$$

$\Sigma_{j=1}^{n} X_j$ in exponential families:

If $X_1, \ldots, X_n$ random sample, $X_j \sim f(x|\theta),\ j = 1, \ldots, n$,

$f(x|\theta) = h(x)c(\theta)\exp(\Sigma_{i=1}^{k} w_i(\theta)t_i(x))$.

Let $T_i(X_1, \ldots, X_n) = \Sigma_{j=1}^{n} t_i(X_j),\ i = 1, \ldots, k$.

If $\{(w_1(\theta), \ldots, w_k(\theta))|\theta \in \Theta\}$ contains an open subset of $\mathbb{R}^k$, $(T_1, \ldots, T_k) \sim f(u_1, \ldots, u_k|\theta)$ where

$f(u_1, \ldots, u_k|\theta) = H(u_1, \ldots, u_k)c(\theta)^n \exp(\Sigma_{i=1}^{k} w_i(\theta)u_i)$.

Example

$X_1, \ldots, X_n$ i.i.d Bernoulli.

$c(p) = 1 - p$, $w(p) = \log \frac{p}{1-p}$, $t_1(x) = x$.

Hence $T_1 = T_1(X_1, \ldots, X_n) = \Sigma_{j=1}^{n} X_j$. This is compatible with

what we know: $\Sigma_{j=1}^{n} X_j \sim \textit{binomial}(n, p)$ which belongs to the exponential family of distributions.

Sampling from the normal distribution

One sample:

Theorem

If $X_1, \ldots, X_n$ is a random sample, $X_j \sim n(\mu, \sigma)$, $j = 1, \ldots, n$

i) $\bar{X} \sim n(\mu, \sigma^2/n)$

ii) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

iii) $\bar{X}$ *and* S^2 *are independent*

Student t: $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}\frac{1}{\sqrt{n-1}}} = \frac{U}{\sqrt{V/p}}$ where

$U \sim n(0,1)$, $V \sim \chi^2_p$, U and V independent and $p = n-1$.
Thus t is a t_p distributed variable.

Density:

Let $U \sim n(0,1)$, $V \sim \chi^2_p$.

$f_{UV}(u,v) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \frac{1}{\Gamma(\frac{p}{2})} \frac{1}{2^{p/2}} v^{\frac{p}{2}-1} e^{-v/2}, \; u, v > 0.$

Then $t = \frac{U}{\sqrt{V/p}}$, $W = V$, with inverses $U = t\sqrt{\frac{W}{p}}$, $V = W$, has density $f_{tW}(t,w) = \frac{1}{\sqrt{2\pi}} e^{-(\frac{t^2}{p}+1)w/2} \frac{1}{\Gamma(\frac{p}{2})} \frac{1}{2^{p/2}} w^{\frac{p+1}{2}-1} \frac{1}{\sqrt{p}}$.

Hence, t has marginal density

$$
\begin{aligned}
f_t(t) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(\frac{t^2}{p}+1)w/2} \frac{1}{\Gamma(\frac{p}{2})} \frac{1}{2^{p/2}} w^{\frac{p+1}{2}-1} \frac{1}{\sqrt{p}} dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{p}} e^{-s} \frac{1}{\Gamma(\frac{p}{2})} \frac{1}{2^{p/2}} [\frac{2s}{\frac{t^2}{p}+1}]^{\frac{p+1}{2}-1} [\frac{2}{\frac{t^2}{p}+1}] ds \\
&= \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{\sqrt{p\pi}} [\frac{1}{\frac{t^2}{p}+1}]^{\frac{p+1}{2}}
\end{aligned}
$$

Not moments of all orders.

$p = 1, (i.e. n = 1)$ Cauchy(0,1).

Two (independent) samples:

$X_1, \ldots, X_n$ is a random sample, $X_j \sim n(\mu_X, \sigma_X^2),\ j = 1, \ldots, n$
$Y_1, \ldots, Y_m$ is a random sample, $Y_j \sim n(\mu_Y, \sigma_Y^2),\ j = 1, \ldots, m$

Let $F = \frac{S_X^2}{\sigma_X^2} / \frac{S_Y^2}{\sigma_Y^2}$.

F is Fisher distributed with n-1 and m-1 degrees of freedom

since $F = \frac{S_X^2}{\sigma_X^2} / \frac{S_Y^2}{\sigma_Y^2} = \frac{[(n-1)\frac{S_X^2}{\sigma_X^2}]/(n-1)}{[(m-1)\frac{S_X^2}{\sigma_X^2}]/(m-1)} \sim \frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)}$.