Transformations of RV's

Suppose X is a discrete RV and $Y = u(X)$ where u is a one-to-one function on $\mathcal{A} \to \mathcal{B}$

Then $f_Y(y) = P(Y=y) = P(X = u^{-1}(y)) = f_X(u^{-1}(y))$

Ex: Invitations to a party are sent to n married couples. Each couple will either both attend wp p or neither will attend wp $(1-p)$, and the couples are mutually indep.

What is the dist'n of the number of people who attend?

X = # couples who attend. $X \sim \text{Bin}(n,p)$. $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$

Y = # people who attend. $Y = 2X$. $X = \frac{1}{2}Y$ $\quad f_Y(y) = f_X\left(\frac{y}{2}\right) = \binom{n}{y/2} p^{y/2} (1-p)^{n-y/2} I_{\{y \in \{0,2,4,\ldots,2n\}\}}$

Suppose X is a continuous RV and $Y = u(X)$, u a one-to-one fn on $\mathcal{A} \to \mathcal{B}$

Then $f_Y(y) = f_X(u^{-1}(y)) \left| \frac{d u^{-1}(y)}{dy} \right|$.

This is the std change of variables from calculus:

$$P(a < Y < b) = P(u^{-1}(a) < X < u^{-1}(b))$$

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

$$= \int_{u^{-1}(a)}^{u^{-1}(b)} f_X(x)\,dx$$

let $y = u(x)$

$x = u^{-1}(y)$

$dx = \frac{d u^{-1}(y)}{dy} dy$

$$= \int_a^b f_X(u^{-1}(y)) \frac{d u^{-1}(y)}{dy}\, dy$$

$$= \int_a^b f_Y(y)\, dy \qquad \text{so} \quad f_Y(y) = f_X(u^{-1}(y)) \frac{d u^{-1}(y)}{dy}$$

$\int_A f_Y(y)\,dy$

Ex: $X \sim \text{Cauchy}(0)$ $\quad Y = \frac{1}{X}$ $\quad$ What is the dist'n of Y?

$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$ $\quad u(x) = \frac{1}{x}$ $\quad u^{-1}(y) = \frac{1}{y}$ $\quad \frac{d u^{-1}(y)}{dy} = -\frac{1}{y^2}$

$f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{1+\frac{1}{y^2}} \cdot \frac{1}{y^2} = \frac{1}{\pi} \cdot \frac{1}{y^2+1}$ $\quad$ so $Y \sim \text{Cauchy}(0)$.

Key Sections in Ch. 4: 4.3, 4.4, 4.6, 4.7, 4.8, 4.10

Suppose X_1 and X_2 have joint pdf $f_{X_1,X_2}(x_1,x_2)$

Let $Y_1 = u_1(X_1,X_2)$, $Y_2 = u_2(X_1,X_2)$, u_1 and u_2 one-to-one. $\qquad x_1 = w_1(y_1,y_2)$

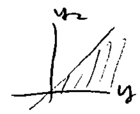
Define the <u>Jacobian</u> of the transformation as $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$ $\qquad \frac{\partial x_i}{\partial y_j} = \frac{\partial w_i(y_1,y_2)}{\partial y_j}$

Then $f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(w_1(y_1,y_2), w_2(y_1,y_2))\,|J|$

Ex: $X_1, X_2 \overset{iid}{\sim} \text{Exp}(\lambda)$ $\qquad$ What is the dist'n of $Y = X_1 + X_2$?

let $u_1(x_1,x_2) = x_1 + x_2$, $u_2 = x_2$ $\qquad w_2(y_1,y_2) = y_2 \quad w_1(y_1,y_2) = y_1 - y_2$

$$J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1 \qquad f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2}$$



$$f_{Y_1,Y_2}(y_1,y_2) = \lambda e^{-\lambda(y_1-y_2)} \lambda e^{-\lambda y_2} (1) = \lambda^2 e^{-\lambda y_1}$$

$\longrightarrow$ does this just factor? no - need indicator fn for range $I\{y_2>0,\, y_1>y_2\}$

$$f_{Y_1}(y_1) = \int_0^{y_1} f_{Y_1,Y_2}(y_1,y_2)\,dy_2 = \int_0^{y_1} \lambda^2 e^{-\lambda y_1}\,dy_2 = \lambda^2 y_1 e^{-\lambda y_1}$$

Thus $Y_1 = X_1 + X_2 \sim \Gamma(2,\lambda)$

In general, if $X_1 \sim \Gamma(\alpha_1,\beta)$, $X_2 \sim \Gamma(\alpha_2,\beta)$ indep, then $X_1 + X_2 \sim \Gamma(\alpha_1+\alpha_2,\beta)$

Other examples in the text:

$X_1, X_2 \overset{iid}{\sim} N(0,1)$. Then $\frac{X_1}{X_2} \sim \text{Cauchy}(0)$

Box-Muller Transformation

If $Y_1, Y_2 \sim \text{Unif}[0,1]$, then $X_1 = (-2\ln Y_1)^{1/2}\cos(2\pi Y_2)$, $X_2 = (-2\ln Y_1)^{1/2}\sin(2\pi Y_2)$ are indep $N(0,1)$.

In general, for $Y = X_1 + X_2$, $X_1 \perp X_2$, $f_Y(y) = \int f_{X_1}(y_1-y_2) f_{X_2}(y_2)\,dy_2$ $\qquad$ convolution formula

Beta

Let $X_1 \sim \Gamma(\alpha_1,1)$, $X_2 \sim \Gamma(\alpha_2,1)$ indep.

Then $\frac{X_1}{X_1+X_2}$ is defined to have a Beta(α_1,α_2) distribution.

Let $Y_1 = X_1 + X_2$, $Y_2 = \frac{X_1}{X_1+X_2}$ $\qquad w_1 = y_1 y_2 \quad w_2 = y_1 - y_1 y_2 \qquad J = \begin{vmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{vmatrix} = -y_1$

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\Gamma(\alpha_1)} (y_1 y_2)^{\alpha_1-1} e^{-y_1 y_2} \frac{1}{\Gamma(\alpha_2)} (y_1(1-y_2))^{\alpha_2-1} e^{-y_1(1-y_2)}\, y_1\, I\{y_1>0,\, 0<y_2<1\}$$

$$= \left(\frac{1}{\Gamma(\alpha_1+\alpha_2)} y_1^{(\alpha_1+\alpha_2)-1} e^{-y_1}\right)\left(\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_2^{\alpha_1-1}(1-y_2)^{\alpha_2-1}\right) I_{\{y_1>0\}} I_{\{0<y_2<1\}}$$

factors $\Rightarrow$ indep. $\qquad Y_1 \sim \Gamma(\alpha_1+\alpha_2, 1)$ as before

$$f_{Y_2}(y_2) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_2^{\alpha_1-1}(1-y_2)^{\alpha_2-1} I_{\{0<y_2<1\}}$$ not symmetric wrt α_1, α_2

some books define $B(\alpha_1,\alpha_2) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$

$E[Y_2] = \frac{\alpha_1}{\alpha_1+\alpha_2} \qquad Var(Y_2) = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2+1)(\alpha_1+\alpha_2)^2} \qquad U(0,1)$ is Beta$(1,1)$

t

Let $W \sim N(0,1)$, $V \sim \chi^2_r$ indep. Then $X = \frac{W}{\sqrt{V/r}} \sim t_r$

$$f_X(x) = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{\pi r}} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}} \qquad x \in \mathbb{R}$$

Note: symmetric

$t_1 \equiv \text{Cauchy}(0)$

$\lim_{r\to\infty} t_r = N(0,1)$

cdf in Table B-IV

Very important for confidence intervals and hypothesis testing.

F

Let $U \sim \chi^2_{r_1}$, $V \sim \chi^2_{r_2}$ indep. Then $X = \frac{U/r_1}{V/r_2} \sim F_{r_1,r_2}$

Note: $\frac{1}{F_{r_1,r_2}} \sim F_{r_2,r_1}$

$(t_r)^2 \sim F_{1,r}$

pdf in text

cdf in Table B-II note two parameters are not exchangeable

Also important for hypothesis testing.

Transformations of n variables

$X_1, \dots, X_n$ have joint pdf $f_{X_1,\dots,X_n}(x_1, \dots, x_n)$

$y_i = u_i(x_1, \dots, x_n)$ for $i = 1, \dots, n$

let $x_i = w_i(y_1, \dots, y_n)$ for $i = 1, \dots, n$

let
$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Then $f_{Y_1,\dots,Y_n}(y_1, \dots, y_n) = f_{X_1,\dots,X_n}(w_1(y_1,\dots,y_n), \dots, w_n(y_1,\dots,y_n))\, |J|$

Sampling Theory

Def: A statistic is any function of observed data (RV's).

Ex: Data: $\{x_1, \dots, x_n\}$

Statistics: $\{x_1, \dots, x_n\}$, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $\min\{x_i\}$, $\max\{x_i\}$, range $= \max\{x_i\} - \min\{x_i\}$

Cannot depend on unknown parameters, but its distribution can.

Ex: $X_1, X_2 \overset{iid}{\sim} \text{Exp}(\lambda)$ with λ unknown

$\bar{x} = \frac{1}{2}(x_1 + x_2) \sim \Gamma(2, \frac{\lambda}{2})$ $\quad \frac{1}{2\lambda}(x_1 + x_2) \sim \Gamma(2, \frac{1}{2})$ but $\frac{1}{2\lambda}(x_1+x_2)$ is not statistic

Def: A random sample is a collection of observations $\{X_1, \dots, X_n\}$ with X_i iid.