Bivariate Random Variables

Median

$$median(F_X) = x\ s.t. F_X(x) = \frac{1}{2}$$

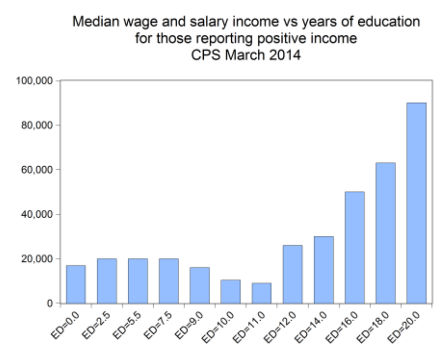
Examples:

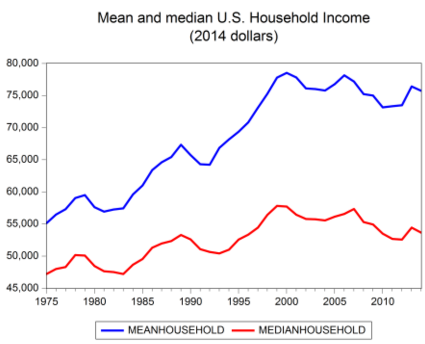
$$x \sim N(\mu, \sigma^2) \Rightarrow median(x) = \mu$$

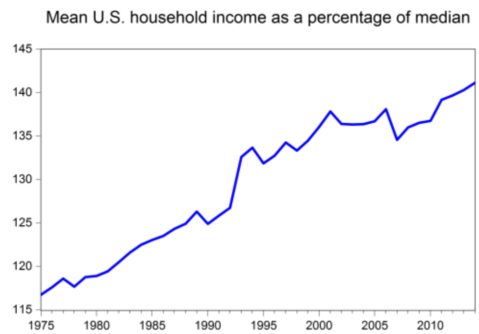
$$x \sim exponential(\lambda) \ \Rightarrow median(x) = \frac{\log(2)}{\lambda}$$

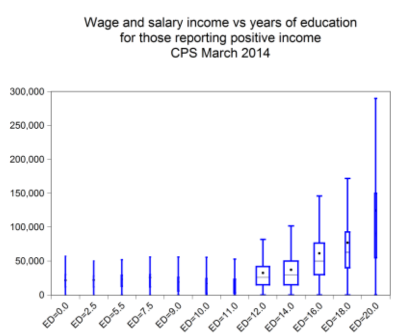
Assignment

- Generate 1,000 samples of 100 iid $\text{exponential}(\lambda)$ variables with $\lambda = 2$, $f(x) = \lambda e^{-\lambda x}, 0 < x < \infty$.
- Plot the distribution of the 25th, 50th, and 75th percentiles of your samples.

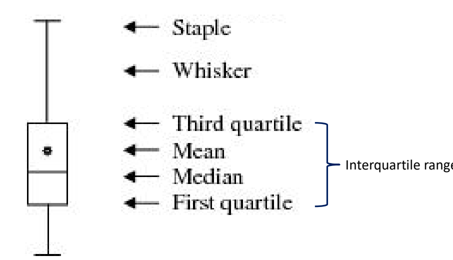








Boxplot

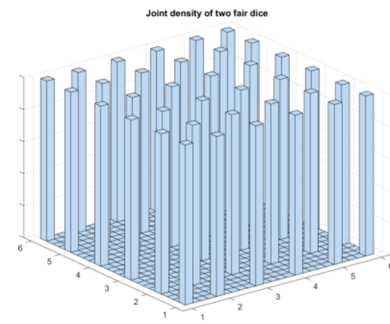


Sometimes width represents proportion of sample, or square root of proportion of sample. Sometimes box is shaded or notched to give confidence intervals for the median

Bivariate pdf

- An *n-dimensional random vector* is a function from a sample space S into $\mathbb{R}^n$.
- A joint pdf (or pmf) maps a random vector into $\mathbb{R}^1$. In particular, a bivariate pdf maps a sample in 2-space (maybe $\mathbb{R}^2$) into $\mathbb{R}$. We have a joint pdf of the form

$$f_{X,Y}(x,y)$$



Moments

$$\mathrm{E}(x) = \iint x f(x,y)\,dx\,dy$$

Etc.

Marginal distributions

Discrete:

$$f_X(x) = \sum_{Y} f_{X,Y}(x,y)$$

Continuous

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

Joint and marginals for two dice

$f(x,y)$	1	2	3	4	5	6	$f(x)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$f(y)$	1/6	1/6	1/6	1/6	1/6	1/6	

Assignment

- Suppose you have a pair of honest dice. What is the probability of a natural (7 or 11)? What is the probability of craps (2, 3, or 12)?

Joint CDF

$$F(x,y) = P(X \le x, Y \le y)$$
$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)dy\,dx$$

- Consider the joint uniform

$$f(x,y) = \begin{cases} 1, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & otherwise \end{cases}$$
$$F(0.1,0.5) = \int_0^{.1} \int_0^{.5} 1dy\,dx = \int_0^{.1} [y]_0^{.5} dx$$
$$= \int_0^{.1} .5\,dx = .05$$

Independent Random Variables

Let (X, Y) be a bivariate random vector with joint pdf $f(x,y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. Then X and Y are *independent random variables* if for all (x, y)

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

Independent Random Variables

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$f_X(x) = \int_{-\infty}^{\infty} [f_X(x) \cdot f_Y(y)]dy$$

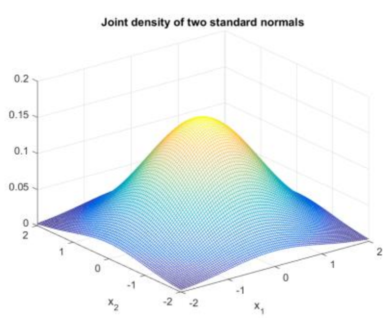
$$= f_X(x) \cdot \int_{-\infty}^{\infty} f_Y(y)dy = f_X(x) \cdot 1$$

Bivariate independent normal pdf

- $x_1 \sim N(0,1)$
- $x_2 \sim N(0,1)$

$$f(x_1, x_2) = \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2}\right] \times \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}\right]$$

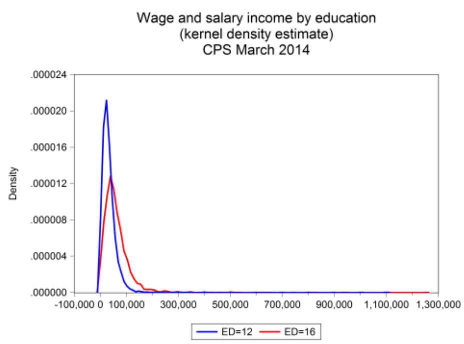
$$= (2\pi)^{-\frac{n=2}{2}} e^{-\frac{1}{2}(x_1^2+x_2^2)}$$



Conditional density

Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the *conditional pdf of Y given $X = x$* is

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$



A conditional density is a density

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

$$= \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f(x,y)dy$$

$$\frac{1}{f_X(x)} \times f_X(x) = 1$$

Bayes theorem again

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

$$f(x|y) = \frac{f(x,y)}{f_y(y)}$$

$$f(y|x) = \frac{f(x|y) \times f_y(y)}{f_X(x)}$$

Law of total probability

$$P(y) = \sum_x P(y|x)P(x)$$

$$f_Y(y) = \int f(y|x)f_X(x)dx$$

Conditional moments

$$\mathrm{E}(y|x) = \int_{-\infty}^{\infty} y \cdot f(y|x)dy$$

$$\mathrm{var}(y|x) = \int_{-\infty}^{\infty} (y - \mathrm{E}(y|x))^2 \cdot f(y|x)dy$$

Independence and conditional distribution

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y)}{f_X(x)} = f_Y(y)$$

Law of Iterated Expectations

Theorem 4.4.3 If X and Y are random variables then

$$\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}(Y)$$

Provided the expectations exist.

Proof

Note that $\mathrm{E}(y|x)$ is a function of x.

$$\mathrm{E}(\mathrm{E}(Y|X))$$
$$= \int \mathrm{E}(y|x)\, f_X(x)dx$$

$$\int \left[\int y \cdot f(y|x)dy \right] f_X(x)dx$$

Substitute in $f(y|x) = f(y,x)/f(x)$

$$= \int \left[\int y \cdot f(y,x)/f(x)dy \right] f_X(x)dx$$

$$= \int \frac{1}{f(x)} \left[\int y \cdot f(y,x)dy \right] f_X(x)dx$$

$$= \int \cancel{\frac{1}{f(x)}} \left[\int y \cdot f(y,x)dy \right] \cancel{f_X(x)}dx$$

$$\mathrm{E}(y) = \int \int y \cdot f(y,x)dy\,dx$$

Law of iterated iterations

$$E(E(y|x_1, x_2)\,|x_1) = E(y|x_1)$$

Hansen calls this "The smaller information set wins."

Forecast errors are mean zero

$$y^e = E(y|x)$$
$$\varepsilon = y - y^e$$
$$E(\varepsilon|x) = E(y|x) - E(E(y|x)\,|x)$$
$$E(y|x) - E(y|x) = 0$$

By the Law of Iterated Expectations $E(\varepsilon) = 0$.

Note: Random walk of stock prices. If theory says $p_t = E(p_{t+1}|I_t)$, then

$$p_{t+1} = E(p_{t+1}|I_t) + \varepsilon$$
$$p_{t+1} = p_t + \varepsilon$$

Dependent Variable: CLOSE
Method: Least Squares
Date: 08/28/16 Time: 11:42
Sample (adjusted): 1/04/1950 7/10/2015
Included observations: 16485 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.059183	0.080990	0.730741	0.4649
CLOSE(-1)	1.000139	0.000113	8887.713	0.0000

R-squared	0.999791	Mean dependent var	472.9167
Adjusted R-squared	0.999791	S.D. dependent var	542.7898
S.E. of regression	7.840210	Akaike info criterion	6.956530
Sum squared resid	1013192.	Schwarz criterion	6.957465
Log likelihood	-57337.19	Hannan-Quinn criter.	6.956838
F-statistic	78991436	Durbin-Watson stat	2.123942
Prob(F-statistic)	0.000000		

$$\log p_{t+1} = \log p_t + \varepsilon_t$$
$$\frac{p_{t+1} - p_t}{p_t} \approx \log p_{t+1} - \log p_t = \varepsilon_t$$

Dependent Variable: LNP
Method: Least Squares
Date: 09/07/17 Time: 12:02
Sample (adjusted): 1/04/1950 7/10/2015
Included observations: 16485 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000555	0.000313	1.775517	0.0758
LNP(-1)	0.999951	5.66E-05	17674.34	0.0000

R-squared	0.999947	Mean dependent var	5.363609
Adjusted R-squared	0.999947	S.D. dependent var	1.336568
S.E. of regression	0.009709	Akaike info criterion	-6.431437
Sum squared resid	1.553717	Schwarz criterion	-6.430501
Log likelihood	53013.12	Hannan-Quinn criter.	-6.431128
F-statistic	3.12E+08	Durbin-Watson stat	1.944971
Prob(F-statistic)	0.000000		

Conditioning theorem

$$\mathrm{E}(x|x) = x$$

$$\mathrm{E}(g(x)|x) = g(x)$$

$$\mathrm{E}(g(x)y|x) = g(x)\,\mathrm{E}(y|x)$$

$$\mathrm{E}(g(x)y) = \mathrm{E}(g(x)\,\mathrm{E}(y|x))$$

Forecast error is uncorrelated with functions of variables used in the forecast

$$\mathrm{E}\big(h(x)(y - E(y|x))\big) = 0$$

Proof:

$$= \mathrm{E}(\mathrm{E}[h(x)(y - E(y|x))|x])$$

By the law of iterated expectations.

$$= \mathrm{E}(h(x)\,\mathrm{E}[(y - E(y|x))|x])$$

By conditioning theorem

$$\mathrm{E}[(y - E(y|x))|x] = 0$$


```
ls log(gdpc96) c log(gdpc96(-1))  log(gdpc96(-2))
fit(d) yfit
show log(gdpc96)-yfit log(gdpc96(-1)) log(gdpc96(-2))
```

Dependent Variable: LOG(GDPC96)
Method: Least Squares
Date: 10/11/15 Time: 11:38
Sample (adjusted): 1947Q3 2015Q2
Included observations: 272 after adjustments

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
C	0.020863	0.007576	2.753776	0.0063
LOG(GDPC96(-1))	1.351018	0.056859	23.76073	0.0000
LOG(GDPC96(-2))	-0.352819	0.056722	-6.220118	0.0000

R-squared	0.999808	Mean dependent var	8.766165
Adjusted R-squared	0.999806	S.D. dependent var	0.634578
S.E. of regression	0.008835	Akaike info criterion	-6.609316
Sum squared resid	0.020995	Schwarz criterion	-6.569546
Log likelihood	901.8670	Hannan-Quinn criter.	-6.593350
F-statistic	698961.9	Durbin-Watson stat	2.061133
Prob(F-statistic)	0.000000		

Covariance Analysis: Ordinary
Date: 10/11/15 Time: 11:42
Sample: 1947Q3 2015Q2
Included observations: 272
Balanced sample (listwise missing value deletion)

Covariance Correlation	LOG(GDPC96)-YFIT...	LOG(GDPC96(-1))	LOG(GDPC96(-2)...
LOG(GDPC96)-YFIT	7.72E-05 1.000000		
LOG(GDPC96(-1))	3.72E-14 6.67E-12	0.403226 1.000000	
LOG(GDPC96(-2))	-5.21E-14 -9.31E-12	0.404156 0.999890	0.405177 1.000000

Prove the following theorem

Theorem: If you want to minimize expected square error given the variable x, conditional expected value is right way to forecast. (Note that $\mathrm{E}(y|x)$ is a function of x.)

$$\mathrm{y^e} = \underset{\mathrm{y^e}}{\operatorname{argmin}} \mathrm{E}\left((\mathrm{y}-\mathrm{y^e})^2\right)$$
$$y^e = \mathrm{E}(y|x)$$

Mean independence

Y is mean-independent of X iff
$$\mathrm{E}(y|x) = \mathrm{E}(y)\ \forall x\ s.t.\ f_X(x) \neq 0$$

- Note: Y and X independent imply mean independence, but the implication does not necessarily follow the other way around. There can be information about higher moments without information about the mean.

Conditional variance identity

Theorem 4.4.7 (Conditional variance identity)
For any two random variables X and Y,
$$\mathrm{var}(X) = \mathrm{E}(var(X|Y)) + \mathrm{var}(\mathrm{E}(X|Y))$$
Provided that the expectations exist.

Covariance

$$\mathrm{cov}(x,y) = \mathrm{E}[(x-E(x))(y-E(y))]$$

$$\mathrm{cov}(x,y) = \iint (x-E(x))(y-E(y))f(x,y)dxdy$$

$$\mathrm{cov}(x,y) = \mathrm{E}(xy) - \mathrm{E}(x)\,\mathrm{E}(y)$$
$$\mathrm{cov}(a+bx, c+dy) = bd \cdot \mathrm{cov}(x,y)$$

Variance rules

$$\mathrm{var}(x+y) = \mathrm{E}([(x-\mathrm{E}(x)) + (y-\mathrm{E}(y))]^2)$$
$$= \mathrm{E}\big((x-\mathrm{E}(x))^2\big) + \mathrm{E}\big((y-\mathrm{E}(y))^2\big)$$
$$+ 2\,\mathrm{E}\big((x-\mathrm{E}(x)) \times (y-\mathrm{E}(y))\big)$$

$$\mathrm{var}(ax+by)$$
$$= a^2\,\mathrm{var}(x) + b^2\,\mathrm{var}(y) + 2ab \cdot \mathrm{cov}(x,y)$$
$$\mathrm{var}(x+y) = \mathrm{var}(x) + \mathrm{var}(y) + 2\,\mathrm{cov}(x,y)$$

Correlation

$$\rho_{x,y} = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)}\,\sqrt{\mathrm{var}(y)}}$$

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$$

Perfect affine relation

$$y = a + bx$$

$$\mathrm{var}(y) = b^2\,\mathrm{var}(x)$$
$$\mathrm{cov}(y,x) = b\,\mathrm{var}(x)$$

$$\rho = \frac{b\,\mathrm{var}(x)}{\sqrt{\mathrm{var}(x)}\,\sqrt{\mathrm{b}^2\mathrm{var}(x)}} = \pm 1$$

Independence and correlation

$$independence \Rightarrow uncorrelated$$

$$mean\ independence \Rightarrow uncorrelated$$

$$uncorrelated \nRightarrow independence$$

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \sim \left(\begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right)$$
$$r_p = \omega r_1 + (1-\omega) r_2$$

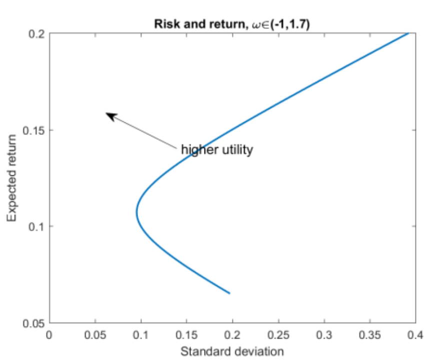
$$r_p \sim (\omega \bar{r}_1 + (1-\omega)\bar{r}_2, \omega^2\sigma_1^2 + (1-\omega)^2\sigma_2^2$$

Copyright Dick Startz 49

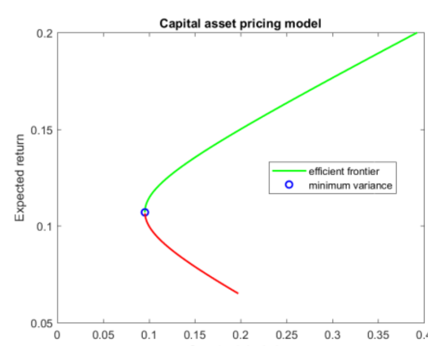
Numerical example

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \sim \left(\begin{bmatrix} .10 \\ .15 \end{bmatrix}, \begin{bmatrix} .1^2 & 0.004 \\ 0.004 & .2^2 \end{bmatrix} \right)$$

Copyright Dick Startz 50



Copyright Dick Startz 51



Copyright Dick Startz 52

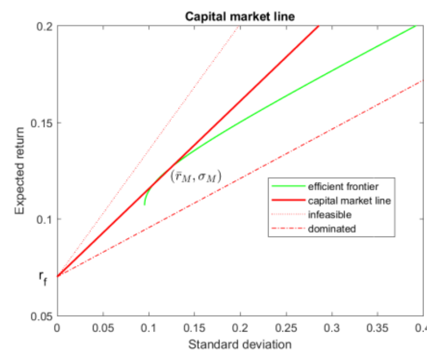
Weight λ in risk-free

- Expected return

$$\lambda r_f + (1-\lambda)\bar{r}_p$$

- Standard deviation

$$(1-\lambda)\sigma_p$$



Capital market line

- Goes through the points

$(r_f, 0)$ and $(\bar{r}_m, \sigma_m)$

- So the equation of the capital market line is

$$r = r_f + \frac{\bar{r}_m - r_f}{\sigma_m}\sigma$$

- Sharpe ratio:

$$\frac{\bar{r}_m - r_f}{\sigma_m}$$

Individual security, weight ω

$$\begin{bmatrix} r_i \\ r_m \end{bmatrix} \sim \left(\begin{bmatrix} \bar{r}_i \\ \bar{r}_m \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & \sigma_{im} \\ \sigma_{im} & \sigma_m^2 \end{bmatrix} \right)$$

$$W \sim \left(\begin{matrix} \omega \bar{r}_i + (1-\omega)\bar{r}_m, \\ \omega^2\sigma_i^2 + (1-\omega)^2\sigma_m^2 + 2\omega(1-\omega)\sigma_{im} \end{matrix} \right)$$

$$W \sim \begin{pmatrix} \omega \bar{r}_i + (1-\omega)\bar{r}_m, \\ \omega^2 \sigma_i^2 + (1-\omega)^2 \sigma_m^2 + 2\omega(1-\omega)\sigma_{im} \end{pmatrix}$$

$$\frac{\partial \mathrm{E}(W)}{\partial \omega} = \bar{r}_i - \bar{r}_m$$

$$\frac{\partial \mathrm{std}(W)}{\partial \omega} = \frac{2\omega\sigma_i^2 - 2(1-\omega)\sigma_m^2 + 2(1-2\omega)\sigma_{im}}{2\mathrm{std}(W)}$$

$$\lim_{\omega \to 0} \frac{\partial \mathrm{std}(W)}{\partial \omega} = \frac{-\sigma_m^2 + \sigma_{im}}{\sigma_m}$$

$$\frac{d\,\mathrm{E}(W)}{d\mathrm{std}(W)} = \frac{\bar{r}_i - \bar{r}_m}{\frac{-\sigma_m^2 + \sigma_{im}}{\sigma_m}} = \frac{\bar{r}_m - r_f}{\sigma_m}$$

$$\frac{d\,\mathrm{E}(W)}{d\mathrm{std}(W)} = \frac{\bar{r}_i - \bar{r}_m}{\frac{-\sigma_m^2 + \sigma_{im}}{\sigma_m}} = \frac{\bar{r}_m - r_f}{\sigma_m}$$

$$\bar{r}_i - \bar{r}_m = \frac{\bar{r}_m - r_f}{\sigma_m} \times \frac{-\sigma_m^2 + \sigma_{im}}{\sigma_m}$$

- Define

$$\beta = \frac{\sigma_{im}}{\sigma_m^2}$$

$$\bar{r}_i = \bar{r}_m + (\bar{r}_m - r_f)(\beta - 1)$$

$$\bar{r}_i - r_f = \beta(\bar{r}_m - r_f)$$

Security market line

- β is the regression coefficient of r_i on r_m.
- The variance of security i is irrelevant. This is much more general than this particular model. Diversifiable risk should not be priced.
- You will also see

$$\bar{r}_i - r_f = \alpha + \beta(\bar{r}_m - r_f)$$

Cauchy-Schwarz Inequality

$$(x \cdot y)^2 \le (x \cdot x)(y \cdot y)$$

where the dots are inner products.
For example

$$|x_1 y_1 + x_2 y_2|^2 \le (x_1^2 + x_2^2)(y_1^2 + y_2^2)$$

$$(\mathrm{E}(XY))^2 \le \mathrm{E}(X^2)\,\mathrm{E}(Y^2)$$

Let

$$X = x - \mu_x, Y = y - \mu_y$$

$$\left(\mathrm{E}\left((x-\mu_x)(y-\mu_y)\right)\right)^2 \le \mathrm{E}\left((x-\mu_x)^2\right) \times \mathrm{E}\left((y-\mu_y)^2\right)$$

$$|\mathrm{cov}(x,y)| \le \sigma_x \sigma_y$$

Minkowski's inequality

Let X and Y be any two random variables. Then for $1 \leq p < \infty$,

$$(\mathrm{E}[|X+Y|^p])^{\frac{1}{p}} \leq (\mathrm{E}[|X|^p])^{\frac{1}{p}} + (\mathrm{E}[|Y|^p])^{\frac{1}{p}}$$

Conditional expectations inequality

For $r \geq 1$,

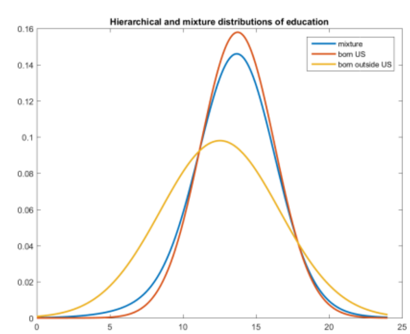
$$\mathrm{E}(|E(y|x)|^r) \leq \mathrm{E}(|y|^r)$$

Hierarchical models and mixture distributions

A random variable X is said to have a mixture distribution if the distribution of X depends on a quantity that also has a distribution.

Example: Mixture of normals

$$f(x) = p \cdot \left[\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}\right] + (1-p) \cdot \left[\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2}\right]$$



Jacobian

f(x): $\mathbb{R}^n \to \mathbb{R}^m$

$$\mathbf{J} = \frac{d\mathbf{f}}{dx} = \begin{bmatrix} \frac{\partial \boldsymbol{f}}{\partial x_1} & \cdots & \frac{\partial \boldsymbol{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

|**J**|

Transformation pdf

- **Theorem 2.1.5** Let X have continuous pdf $f_X(x)$ on $\mathcal{X}$ and let $Y = g(X)$, where $g(\cdot)$ is a monotone function and that $g^{-1}(\cdot)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|, y \in \mathcal{Y} \\ 0, \quad \text{otherwise} \end{cases}$$

Bivariate transformation

Let (X, Y) be a bivariate random vector with joint pdf $f_{X,Y}(x, y)$ and consider the one-to-one functions

$$u = g_u(x, y)$$
$$v = g_v(x, y)$$

With inverse functions

$$x = h_x(u, v)$$
$$y = h_y(u, v)$$

Change of variables

$$\mathbf{J} = \left\| \begin{bmatrix} \frac{\partial h_x}{\partial u} & \frac{\partial h_x}{\partial v} \\ \frac{\partial h_y}{\partial u} & \frac{\partial h_y}{\partial v} \end{bmatrix} \right\|$$

$$f_{U,V}(u, v) = f_{X,Y}\left(h_x(u, v), h_y(u, v)\right)|\mathbf{J}|$$

where |**J**| is the absolute value of the determinant.

Sum of two independent $U(0,1)$

$$f_{X,Y} = \begin{cases} 1, 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, otherwise \end{cases}$$

Transformations

$$u = x + y$$
$$v = x - y$$

The inverse functions are given by

$$x = h_x(u, v) = \frac{u+v}{2}$$
$$y = h_y(x, y) = \frac{u-v}{2}$$
$$\mathbf{J} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

Joint density

$$0 \leq u \leq 2$$
$$-1 \leq v \leq 1$$
$$f_{x,y}\left(h_x(u, v), h_y(u, v)\right)$$
$$= \begin{cases} 1, 0 \leq u \leq 2, -1 \leq v \leq 1 \\ 0, otherwise \end{cases}$$
$$f_{u,v}(u, v) = \begin{cases} 1 \times \left|-\frac{1}{2}\right|, 0 \leq u \leq 2, -1 \leq v \leq 1 \\ 0, otherwise \end{cases}$$

Marginal distribution of the sum

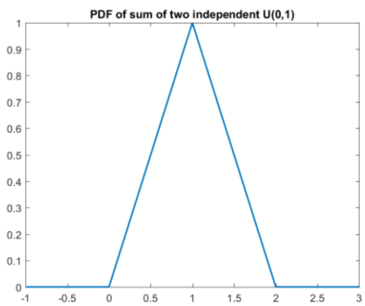
$$f_U = \int_{-\infty}^{\infty} f_{U,V}(u, v)dv$$

Suppose $0 \leq u \leq 1$, then $-u \leq v \leq u$.

$$f_U = \int_{-u}^{u} \frac{1}{2} dv = \frac{1}{2}[v]_{-u}^{u} = u$$

Suppose $1 \leq u \leq 2$, then $-(2-u) \leq v \leq (2-u)$

$$f_U = \int_{-(2-u)}^{(2-u)} \frac{1}{2} dv = \frac{1}{2}[v]_{-(2-u)}^{(2-u)} = 2 - u$$



Marginal distribution of the difference

$$f_V = \int_{-\infty}^{\infty} f_{U,V}(u,v)\,du$$

If $-1 < v < 0$, then let's look at $-2v < u < 2$

$$f_V(v) = \left[\frac{1}{2}u\right]_{-2v}^{2} = 1 + v$$

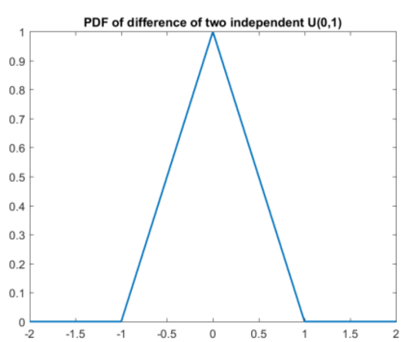
If $0 < v < 1$, then let's look at $2v < u < 2$

$$f_V(v) = \left[\frac{1}{2}u\right]_{2v}^{2} = 1 - v$$

Now consider

$$f_{u,v}(1,1) = \frac{1}{2} \neq f_u(1) \times f_v(1) = 1 \times 0 = 0$$

Therefore u, v are not independent.



Independence of sum and difference of independent uniforms

Suppose $0 \leq u \leq 1$, then $-u \leq v \leq u$.

$$f_U = \int_{-u}^{u} \frac{1}{2}\,dv = \frac{1}{2}[v]_{-u}^{u} = u$$

If $0 < v < 1$, then let's look at $2v < u < 2$

$$f_V(v) = \left[\frac{1}{2}u\right]_{2v}^{2} = 1 - v$$

Now consider

$$f_{u,v}(1,1) = \frac{1}{2} \neq f_u(1) \times f_v(1) = 1 \times 0 = 0$$

Therefore u, v are not independent.

Sum and difference of independent, standard normals

$$f(x,y) = (2\pi)^{-1} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2}$$

Transformations

$$u = x + y$$
$$v = x - y$$

Change of variables

$$f_{U,V}(u,v) = (2\pi)^{-1} e^{-\frac{1}{2}\left(\frac{u+v}{2}\right)^2} e^{-\frac{1}{2}\left(\frac{u-v}{2}\right)^2} \times \left|-\frac{1}{2}\right|$$

$$f_{U,V}(u,v) = (2\pi)^{-1} e^{-\frac{1}{2}\left[\left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2\right]} \times \left|-\frac{1}{2}\right|$$

Adding exponents

$$-\frac{1}{2}\left(\frac{1}{2}\right)^2 [(u+v)^2 + (u-v)^2]$$

$$= -\frac{1}{2}\left(\frac{1}{2}\right)^2 [(u^2 + v^2 + 2uv) + (u^2 + v^2 -$$

Joint density

$$f_{U,V}(u,v) = \left(2\pi\sqrt{2}^2\right)^{-1} e^{-\frac{1}{2}\left(\frac{u}{\sqrt{2}}\right)^2} e^{-\frac{1}{2}\left(\frac{v}{\sqrt{2}}\right)^2}$$
$$u, v \sim i.i.d.\, N\left(0, \sqrt{2}^2\right)$$

Quick check:

$$\mathrm{var}(x+y) = \mathrm{var}(x) + \mathrm{var}(y) + 2 \times 0$$
$$1 + 1 = 2$$

Functions of independent random variables

Let X and Y be independent random variables.

If

$$U = g(X)$$
$$V = h(Y)$$

Then U and V are independent.

Convolution formula

Theorem 5.2.9 If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw$$

Note: If the limits of integration may be modified if either pdf has a limited support.

Note: Because the sum is symmetric in X and Y, the convolution can also be written

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-w) f_Y(w) dw$$

Sum of independent $x \sim U(0,1)$ and $y \sim U(0,1)$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\, dw$$

$$f_Z(z) = \int_{0}^{1} 1 \times f_Y(z-w)\, dw$$

$$f_Z(z) = \int_{0}^{1} f_Y(z-w)\, dw$$

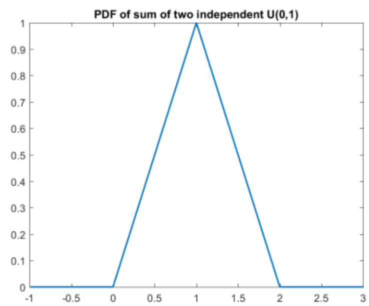
The integrand is zero except if $0 \le z - w \le 1$, in which case it is 1. So if $0 \le z \le 1$ w can take on any value up to z

$$f_Z(z) = \int_{0}^{z} 1\, dw = z$$

If z is between 1 and 2 w can be as low as $z - 1$ and as high as 1.

$$f_Z(z) = \int_{z-1}^{1} 1\, dw = 2 - z$$

Otherwise the density equals 0



Sum of independent standard normals

$$f_Z(z) = \int_{-\infty}^{\infty} \phi(w)\phi(z-w)dw$$

We can write the innards as

$$\frac{1}{\sqrt{2\pi}}\exp[-\frac{w^2}{2}] \times \frac{1}{\sqrt{2\pi}}\exp[-\frac{(w-z)^2}{2}]$$

$$= \frac{1}{2\pi} \times \exp\left(-\frac{z^2}{4}\right)\exp\left(-\left(w-\frac{z}{2}\right)^2\right)$$

Sum of independent standard normals

Putting this back into the integral gives

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \times \exp\left(-\frac{z^2}{4}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \times \exp\left(-\left(w-\frac{z}{2}\right)^2\right) dw$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \times \exp\left(-\frac{z^2}{4}\right) \sqrt{.5} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{.5}} \times \exp\left(-\frac{1}{2\cdot .5}\left(w-\frac{z}{2}\right)^2\right) dw$$

$$w \sim N\left(\frac{z}{2}, \left(\sqrt{.5}^2\right)\right)$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \times \exp\left(-\frac{z^2}{4}\right)\sqrt{.5} = \frac{1}{\sqrt{2\pi\sqrt{2}^2}}\exp\left(\frac{1}{2\times\sqrt{2}^2}z^2\right) \sim N(0,2)$$

Bivariate normal

If X and Y are joint normal, with means μ_x and μ_y, variances σ_x^2 and σ_y^2, and correlation ρ, we can write the joint density as

$$f(x,y) = \left(2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}\right)^{-1}$$
$$\times\, e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right)}$$

Nice properties

a) The marginal distribution of X is $N(\mu_x, \sigma_x^2)$.
b) The marginal distribution of Y is $N\left(\mu_y, \sigma_y^2\right)$.
c) For any constants a and b, the distribution of $ax + by$ is $N\left(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2 +\right.$

Uncorrelated normals are independent

$$f(x,y) = \left(2\pi\sigma_x\sigma_y\cancel{\sqrt{1-\rho^2}}\right)^{-1}$$
$$\times\, e^{-\frac{1}{2\cancel{(1-\rho^2)}}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \cancel{-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)}\right)}$$

Variance-covariance matrix

Covariance

$$\sigma_{xy} = \rho\sigma_x\sigma_y$$

Variance-covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

$$f(x,y) = (2\pi)^{-\frac{2}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\Sigma^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right)$$

Conditional normals

- $y|x \sim N,\ x|y \sim N$

$$\mathrm{E}(y|x) = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x)$$
$$\mathrm{E}(x|y) = \mu_x + \frac{\sigma_{xy}}{\sigma_y^2}(y - \mu_y)$$
$$\mathrm{var}(y|x) = \sigma_y^2(1-\rho^2)$$
$$\mathrm{var}(x|y) = \sigma_x^2(1-\rho^2)$$
$$\frac{\mathrm{var}\,\mathrm{E}(y|x)}{\mathrm{var}(y)} = \rho^2$$

Conditional normals

- Theory

$$\mathrm{E}(y|x) = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x)$$

- Regression

$$\hat{\beta} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$\frac{\mathrm{var}\ \mathrm{E}(y|x)}{\mathrm{var}(y)} = \rho^2$ regression analogy

$$y = \beta x + \varepsilon$$

The forecast is

$$\mathrm{E}(y|x) = \beta x$$

The explained variation is

$$\mathrm{var}(\mathrm{E}(y|x))$$

So the R^2 is

$$R^2 = \frac{\mathrm{var}(\mathrm{E}(y|x))}{\mathrm{var}(y)}$$

Question for today

Suppose that employers observe a signal of productivity, mp_i, with error.

$$s_i = mp_i + \varepsilon_i$$

where marginal product and the error are joint normal, and uncorrelated, and the mean error is zero. So the error is pure noise. If employers observe the signal for individuals and pay expected marginal product, what does the wage schedule look like, and in particular what happens to an individual's wage if they increase their productivity by 1.0?

Simple lemons

Suppose used cars have a value distribution $f(v)$. You know the value of your car, v_i, but no one else does. For a cost c you can have the value of your car creditably certified by a mechanic. If you do, the car will sell for v. If not the car will sell for the average value of uncertified cars.

Marginal lemon

- Suppose we call the value at which certification is marginal v^*. Then uncertified cars sell for

$$\mathrm{E}(v|v < v^*)$$

Or

$$\int_{-\infty}^{v^*} v f(v|v \leq v^*) dv = \int_{-\infty}^{v^*} \frac{v f(v)}{\int_{-\infty}^{v^*} f(v) dv} dv$$

Gain from certification

$$v_i - \int_{-\infty}^{v^*} \frac{v f(v)}{\int_{-\infty}^{v^*} f(v) dv} dv - c$$

$$0 = v^* - \int_{-\infty}^{v^*} \frac{v f(v)}{\int_{-\infty}^{v^*} f(v) dv} dv - c$$

Uniform example

$$v \sim U[0, b]$$
$$f(v) = \frac{1}{b}$$
$$\int_{-\infty}^{v^*} f(v) dv = F(v^*) = \frac{1}{b} [v]_0^{v^*} = \frac{v^*}{b}$$

So conditional expectation is:

$$\frac{1}{(v^*/b)} \int_{-\infty}^{v^*} v f(v) dv = \frac{1}{(v^*/b)} \frac{1}{2b} [v^2]_0^{v^*} = \frac{{v^*}^2}{2v^*} = \frac{v^*}{2}$$

Uniform example

$$0 = v^* - \frac{v^*}{2} - c$$
$$v^* = 2c$$

Multivariate normal

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \\ \vdots & \vdots & \ddots & \\ \sigma_{1n} & & & \sigma_n^2 \end{bmatrix}\right)$$

$$f(x) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

Marginal and conditional distributions

$$\vec{x}_1 \sim N(\vec{\mu}_1, \Sigma_1)$$

$$\mathrm{E}(\vec{x}_1|\vec{x}_2) = \vec{\mu}_1 + \Sigma_{12}\Sigma_2^{-1}(\vec{x}_2 - \vec{\mu}_2)$$

Conditional distribution and regression

- Theory

$$\mathrm{E}(\vec{x}_1|\vec{x}_2) = \vec{\mu}_1 + \Sigma_{12}\Sigma_2^{-1}(\vec{x}_2 - \vec{\mu}_2)$$

- Regression

$$\hat{\beta} = (X'X)^{-1}X'y$$

Linear combinations of joint normals

If L is an $m \times k$ matrix of constants then

$$Lx \sim N(L\mu, L\Sigma L')$$

Useful application of linear combination of normals

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 \mathrm{I})$$

where X is an $n \times k$ fixed matrix.

$$\begin{aligned}\hat{\beta} &\equiv (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + (X'X)^{-1}X'\varepsilon\end{aligned}$$

$$\begin{aligned}&\hat{\beta} \sim N(\beta, (X'X)^{-1}X'\sigma^2 \mathrm{I} X(X'X)^{-1}) \\ &= N(\beta, \sigma^2(X'X)^{-1})\end{aligned}$$

Vector bivariate normal

If

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{bmatrix}\right)$$

Then x and y are independent iff $\Sigma_{xy} = 0$.

Linear combinations of x and y can be written

$$\begin{bmatrix} L_x & 0 \\ 0 & L_y \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} L_x\mu_x \\ L_y\mu_y \end{bmatrix}, \begin{bmatrix} L_x\Sigma_{xx}L_x{}' & L_x\Sigma_{xy}L_y' \\ L_x\Sigma_{xy}L_y' & L_y\Sigma_{yy}L_y' \end{bmatrix}\right)$$

$L_x x$ and $L_y y$ are independent iff their covariance, $L_x\Sigma_{xy}L_y'$, equals zero.

Scalar Cholesky factorization

Cholesky factorization of σ^2 is σ. If

$$\varepsilon \sim N(0,1)$$

then

$$\begin{gathered} x = \sigma\varepsilon \sim N(0, \sigma^2) \\ x \times \sigma^{-1} \sim N(0,1) \end{gathered}$$

Cholesky factorization

$\exists C$ s.t. C is lower triangular, nonsingular.

$$CC' = \Sigma$$

$$C = \begin{bmatrix} \sigma_1 & 0 \\ \dfrac{\sigma_{12}}{\sigma_1} & \sigma_2^2 - \left(\dfrac{\sigma_{12}}{\sigma_1}\right)^2 \end{bmatrix}$$

n.b. Matlab does upper triangular, $C'C = \Sigma$

If ε is distributed i.i.d. standard normal, $\varepsilon \sim N(0, I)$,

$$C\varepsilon \sim N(C \times 0, CIC') = N(0, \Sigma)$$

If $x \sim N(0, \Sigma)$,

$$C^{-1}x \sim N(0, C^{-1}\Sigma C'^{-1}) = N(0, C^{-1}[CC']C'^{-1})$$
$$= N(0, [C^{-1}C][C'C'^{-1}]) = N(0, I)$$

Example

If $\varepsilon_1, \varepsilon_2$ are i.i.d standard normal, and

$$x_1 = \sigma_1 \varepsilon_1$$
$$x_2 = \frac{\sigma_{12}}{\sigma_1}\varepsilon_1 + \left(\sigma_2^2 - \left(\frac{\sigma_{12}}{\sigma_1}\right)^2\right)\varepsilon_2$$

then

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

χ^2

Theorem: The sum of the squares of n independent standard normals is distributed χ_n^2.

Examples:

If $\varepsilon \sim iidN(0,1)$, then $\sum \varepsilon^2 \sim \chi_n^2$.

If $x_i \sim ind\ N(\mu_i, \sigma_i^2)$, then $\sum \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_n^2$.

χ_d^2 or $\chi^2(d)$

$$x \sim \chi_d^2$$
$$f(x) = \frac{1}{2^{\frac{d}{2}}\Gamma\left(\frac{d}{2}\right)} x^{\frac{d}{2}-1} e^{-\frac{x}{2}}$$
$$\mathrm{E}(x) = d$$
$$\mathrm{var}(x) = 2d$$

- If $\varepsilon \sim N(0,1)$, $\mathrm{E}(\varepsilon^2) = 1$

Class assignment

Prove that if $x \sim \chi_1^2$,$\sqrt{x}$ is distributed standard half-normal, that is, like a standard normal distribution defined only on a non-negative support.

Sum of independent χ^2 is χ^2

If $x_i \sim ind \chi_{k_i}^2$, then

$$x_1 + \cdots + x_k \sim\sim \chi_{\sum k_i}^2$$

Example

$$y_i, i = 1, \ldots, n \sim N\left(\mu_y, \sigma^2\right)$$

μ_y is known

$$H_0\colon \sigma^2 = \sigma_0^2$$

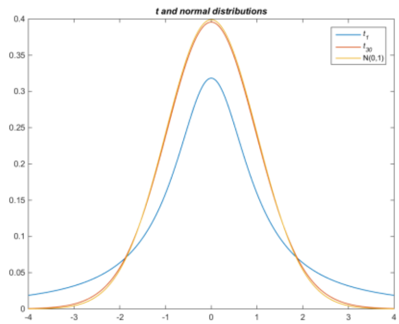
$$\sum \left(\frac{y_i - \mu_y}{\sigma_0}\right)^2 \sim \chi_n^2$$

t-distribution

If $x \sim N(0,1)$ and $S \sim \chi_d^2$ and x, S are independent

$$\frac{x}{\sqrt{S/d}} \sim t_d$$

$$\lim_{n \to \infty} t_n = N(0,1)$$



F-distribution

If $s_1 \sim \chi^2_{d_1}$ and $s_2 \sim \chi^2_{d_2}$, s_1, s_2 independent

$$\frac{s_1/d_1}{s_2/d_2} \sim F(d_1, d_2)$$

$$t_d^2 = F(1, d)$$
$$\lim_{d_2 \to \infty} F(d_1, d_2) = \chi^2_{d_1}/d_1$$

