Econometricks: Short guides to econometrics

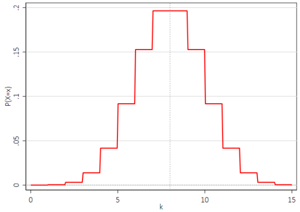
Trick 01: Review of Probability Theory

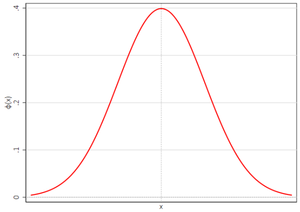
Davud Rostam-Afschar (Uni Mannheim)

Content

Discrete and continuous random variables

- A random variable X is **discrete** if the set of outcomes x is either finite or countably infinite.
- The random variable X is **continuous** if the set of outcomes x is infinitely divisible and, hence, not countable.





Discrete probabilities

For values x of a discrete random variable X,
the **probability mass function** (pmf)

$$f(x) = Prob(X = x).$$

The axioms of probability require

$$0 \leq Prob(X = x) \leq 1,$$

$$\sum_x f(x) = 1.$$

Discrete cumulative probabilities

For values x of a discrete random variable X,
the **cumulative distribution function**

$$F(x) = \sum_{X \leq x} f(x) = Prob(X \leq x),$$

where

$$f(x_i) = F(x_i) - F(x_{i-1}).$$

Example

Roll of a six-sided die

x	$f(x)$	$F(X \leq x)$
1	$f(1) = 1/6$	$F(X \leq 1) = 1/6$
2	$f(2) = 1/6$	$F(X \leq 2) = 2/6$
3	$f(3) = 1/6$	$F(X \leq 3) = 3/6$
4	$f(4) = 1/6$	$F(X \leq 4) = 4/6$
5	$f(5) = 1/6$	$F(X \leq 5) = 5/6$
6	$f(6) = 1/6$	$F(X \leq 6) = 6/6$

What's the probability that you roll a 5 or higher?
$F(X \geq 5) = 1 - F(X \leq 4) = 1 - 2/3 = 1/3$.

Continuous probabilities

For values x of a continuous random variable X, the probability is zero but the area under $f(x) \geq 0$ in the range form a to b is the **probability density function** (pdf)

$$Prob(a \leq x \leq b) = Prob(a < x < b) = \int_a^b f(x)dx \geq 0.$$

The axioms of probability require

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

$f(x) = 0$ outside the range of x.
The **cumulative distribution function** (cdf) is

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

$$f(x) = \frac{dF(x)}{dx}.$$

Cumulative distribution function

For continuous and discrete variables, $F(x)$ satisfies

Definition

Properties of cdf.

- $0 \leq F(x) \leq 1$
- *If* $x > y$, *then* $F(x) \geq F(y)$
- $F(+\infty) = 1$
- $F(-\infty) = 0$

and

$$Prob(a < x \leq b) = F(b) - F(a).$$

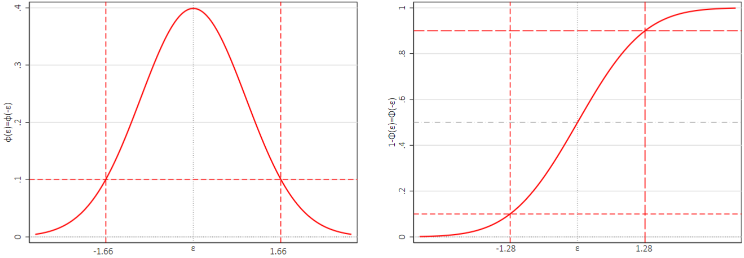
Symmetric distributions

For symmetric distributions

$$f(\mu - x) = f(\mu + x)$$

and

$$1 - F(x) = F(-x).$$



Mean of a random variable

The **mean**, or **expected value**, of a discrete random variable is

$$\mu = E[x] = \sum_x x f(x) \tag{1}$$

Example

Roll of a six-sided die

x	$f(x) = 1/n$	$F(X \leq x) = (x - a + 1)/n$
a = 1	$f(1) = 1/6$	$F(X \leq 1) = 1/6$
2	$f(2) = 1/6$	$F(X \leq 2) = 2/6$
3	$f(3) = 1/6$	$F(X \leq 3) = 3/6$
4	$f(4) = 1/6$	$F(X \leq 4) = 4/6$
5	$f(5) = 1/6$	$F(X \leq 5) = 5/6$
b = 6	$f(6) = 1/6$	$F(X \leq 6) = 6/6$

What's the expected value from rolling the dice?
$E[x] = 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6 = 3.5$.
This is the mean (and the median) of a uniform distribution $(n + 1)/2 = (a + b)/2 = 3.5$.

Mean of a random variable

For a continuous random variable x, the expected value is

$$E[x] = \int_x x f(x) dx.$$

Example

The continuous uniform distribution is $1/(b-a)$ for $a \leq x \leq b$ and 0 otherwise.

$$E[x] = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx.$$

Antiderivative of x is $x^2/2$

$$E[x] = \frac{1}{b-a} b^2/2 - a^2/2 = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

The mean (and the median) is again $(a+b)/2 = 3.5$.

For a function $g(x)$ of x, the expected value is $E[g(x)] = \sum_x g(x) Prob(X = x)$ or $E[g(x)] = \int_x g(x) f(x) dx$. If $g(x) = a + bx$ for constants a and b, then $E[a + bx] = a + bE[x]$.

Variance of a random variable

The **variance** of a random variable $\sigma^2 > 0$ is

$$\sigma^2 = Var[x] = E[(x-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{if } x \text{ is discrete,} \\ \\ \int_x (x-\mu)^2 f(x)dx & \text{if } x \text{ is continuous.} \end{cases} \tag{2}$$

Example

Roll of a six-sided die. What's the variance $V[x]$ from rolling the dice?
The probability of observing x, $Pr(X=x) = 1/n$, is discretely uniformly distributed

$$E[x] = \frac{n+1}{2};\ (E[x])^2 = \frac{(n+1)^2}{4}.$$

$$E[x^2] = \sum_x Pr(X=x) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{(n+1)(2n+1)}{6}$$ due to the sequence sum of squares.

$$V[x] = E[x^2] - (E[x])^2.$$

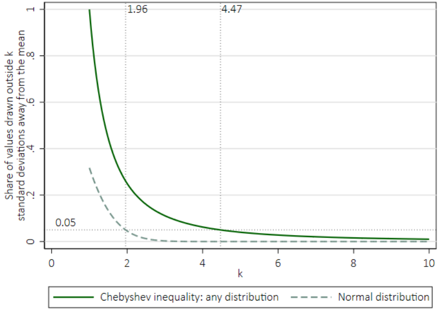
$V[x] = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12} = (6^2-1)/12 \approx 2.92.$

Chebychev inequality

For any random variable x and any positive constant k,

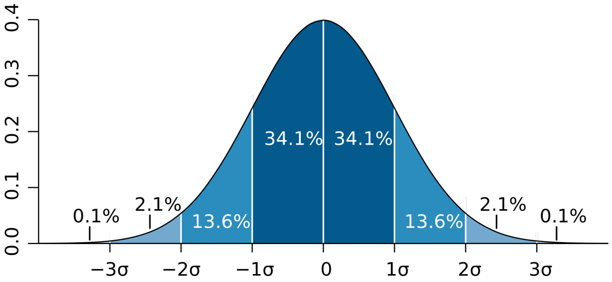
$$\Pr(\mu - k\sigma \leq x \leq \mu + k\sigma) \geq \frac{1}{k^2}.$$

Share outside k standard deviations.
If x is normally distributed, the bound is $1 - (2\Phi(k) - 1)$.



95% of the observations are within 1.96 standard deviations for normally distributed x. If x is not normal, 95% are at most within 4.47 standard deviations.

Normal coverage



Central moments of a random variable

The central moments are

$$\mu_r = E[(x - \mu)^r].$$

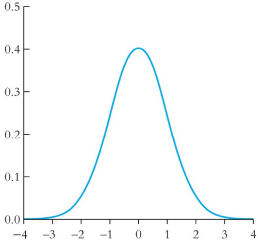
Example

Moments. Two measures often used to describe a probability distribution are

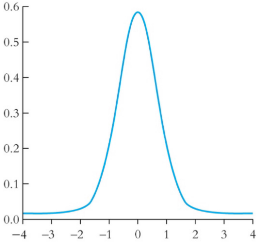
- expectation $= E[(x - \mu)^1]$
- variance $= E[(x - \mu)^2]$
- skewness $= E[(x - \mu)^3]$
- kurtosis $= E[(x - \mu)^4]$

The skewness is zero for symmetric distributions.

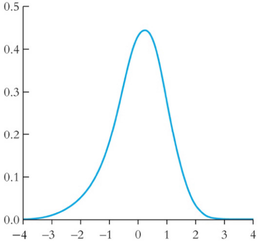
Higher order moments



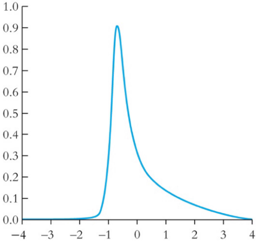
(a) Skewness = 0, kurtosis = 3



(b) Skewness = 0, kurtosis = 20



(c) Skewness = −0.1, kurtosis = 5



(d) Skewness = 0.6, kurtosis = 5

Moment generating function

For the random variable X, with probability density function $f(x)$, if the function

$$M(t) = E[e^{tx}].$$

exists, then it is the **moment generating function**(MGF).

- Often simpler alternative to working directly with probability density functions or cumulative distribution functions
- Not all random variables have moment-generating functions

The nth moment is the nth derivative of the moment-generating function, evaluated at $t = 0$.

Example

The MGF for the standard normal distribution with $\mu = 0, \sigma = 1$ is

$$M_z(t) = e^{\mu t + \sigma^2 t^2/2} = e^{t^2/2}.$$

If x and y are independent, then the MGF of $x + y$ is $M_x(t)M_y(t)$.

Moment generating function

For $x \sim N(\mu, \sigma^2)$ for some $\mu, \sigma > 0$ with moment generating function $M_x{}'(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$, the first moment generating function of x is

$$E[(x-\mu)^1] = M_x{}'(t) = (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Example

$$E[(x-\mu)^1] = M_x{}'(t) = \frac{d\left[\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\right]}{dt}$$

$$= \frac{d\left[\mu t + \frac{1}{2}\sigma^2 t^2\right]}{dt} \frac{d\left[\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\right]}{d(\mu t + \frac{1}{2}\sigma^2 t^2)}$$

$$= (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Moment generating function

If $x \sim N(0,1)$,

- the skewness is $E[(x-\mu)^3] = 0$ and
- the kurtosis is $E[(x-\mu)^4] = 3$.

Example

$$E[(x-\mu)^1] = M_x{}'(t) = (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \text{ with } \mu = 0, \sigma = 1, t = 0 : E[x] = \mu = 0$$

$$E[(x-\mu)^2] = M_x{}''(t) = \left(\sigma^2 + (\mu + \sigma^2 t)^2\right)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$\text{with } \mu = 0, \sigma = 1, t = 0 : E[(x-\mu)^2] = \sigma^2 = 1$$

$$E[(x-\mu)^3] = M_x{}'''(t) = \left(3\sigma^2(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^3\right)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$\text{with } \mu = 0, \sigma = 1, t = 0 : E[(x-\mu)^3] = 0$$

$$E[(x-\mu)^4] = M_x{}^{(4)}(t) = \left(3\sigma^4 + 6\sigma^2(\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4\right)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$\text{with } \mu = 0, \sigma = 1, t = 0 : E[(x-\mu)^4] = 3.$$

Approximating mean and variance

For any two functions $g_1(x)$ and $g_2(x)$,

$$E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]. \tag{3}$$

For the general case of a possibly nonlinear $g(x)$,

$$E[g(x)] = \int_x g(x)f(x)dx, \tag{4}$$

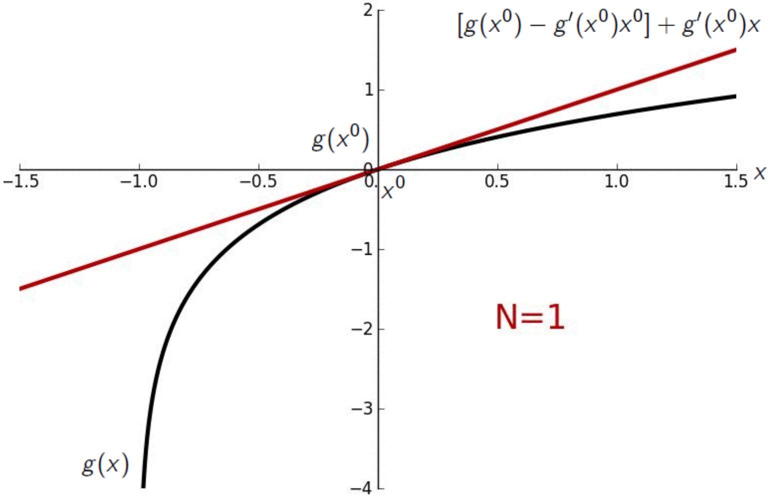
and

$$Var[g(x)] = \int_x (g(x) - E[g(x)])^2 f(x)dx. \tag{5}$$

$E[g(x)]$ and $Var[g(x)]$ can be approximated by a first order linear Taylor series:

$$g(x) \approx [g(x^0) - g'(x^0)x^0] + g'(x^0)x. \tag{6}$$

Taylor approximation Order 1



Thanks to Shiyu Chang

Approximating mean and variance

A natural choice for the expansion point is $x^0 = \mu = E(x)$. Inserting this value in Eq. (6) gives

$$g(x) \approx [g(\mu) - g'(\mu)\mu] + g'(\mu)x, \tag{7}$$

so that

$$E[g(x)] \approx g(\mu), \tag{8}$$

and

$$Var[g(x)] \approx [g'(\mu)]^2 Var[x]. \tag{9}$$

Example

Isoelastic utility. $c_{bad} = 10.00$ Euro; $c_{good} = 100.00$ Euro; probability good outcome 50%

$\mu = E[c] = 1/2 \times c_{bad} + 1/2 \times c_{good} =$55.00 Euro

$$u(c) = c^{1/2}$$

$u(\mu) = 7.42$ approximates $E[u(c)] = 1/2 \times 10^{1/2} + 1/2 \times 100^{1/2} = 6.58$

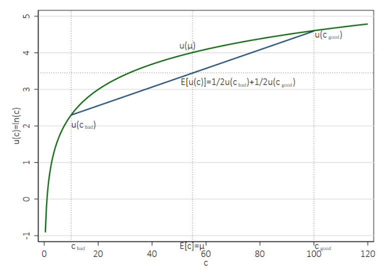
Approximating mean and variance

Example

Isoelastic utility. $c_{bad} = 10.00$ Euro; $c_{good} = 100.00$ Euro; probability good outcome 50%; $\mu = 55.00$ Euro

$$u(c) = \ln(c)$$

$u(\mu) = 4.01$ approx. $E[u(c)] = 1/2 \times \ln(10) + 1/2 \times \ln(100) = 3.45$



Jensen's inequality:
$E[g(x)] \leq g(E[x])$ if $g''(x) < 0$.

$V[u(c)] \approx (1/55)^2((10-55)^2 + (100-55)^2) = 1.34$
$V[u(c)] = (\ln(10) - E[u(c)])^2 + (\ln(100) - E[u(c)])^2 = 2.65$

Useful rules

- $Var[x] = E[x^2] - \mu^2$
- $E[x^2] = \sigma^2 + \mu^2$
- If a and b constants, $Var[a + bx] = b^2 Var[x]$
- $Var[a] = 0$
- If $g(x) = a + bx$ and a and b are constants, $E[a + bx] = a + bE[x]$
- Coverage $\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$
- Skewness $= E[(x - \mu)^3]$
- Kurtosis $= E[(x - \mu)^4]$
- For symmetric distributions $f(\mu - x) = f(\mu + x)$; $1 - F(x) = F(-x)$
- $E[g(x)] \approx g(\mu)$
- $Var[g(x)] \approx [g'(\mu)]^2 Var[x]$

References

Greene, W. H. (2011): *Econometric Analysis*. Prentice Hall, 5 edn.

Pishro-Nik, H. (2014): *Introduction to Probability, Statistics, and Random Processes*. Kappa Research LLC.