Common Distributions

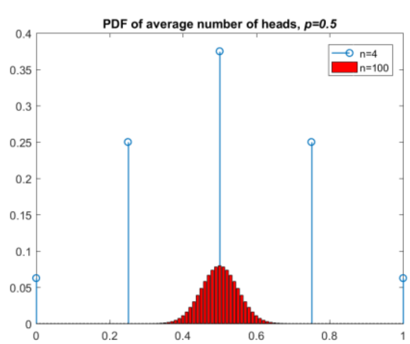
Binomial distribution

$$f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\mathrm{E}(x) = np$$
$$\mathrm{var}(x) = np(1-p)$$

Binomial divided by number of trials:

$$\mathrm{E}(x/n) = p$$
$$\mathrm{var}(x/n) = p(1-p)/n$$



Write a Matlab program that plots two figures (one figure for small n and one for a large n), each comparing two cdfs. For both figures, one cdf is the number of heads out of n coin flips with a probability of heads equals to $p = 1/2$. The other cdf is for a normal distribution with mean np and standard deviation $\sqrt{n/4}$. The Matlab function binocdf will give you the former and normcdf will give you the latter.

Average of standard normals

$$x \sim N(0,1)$$
$$\bar{x} \sim N\left(0, {}^{1}/_{n}\right)$$

$$p(\sqrt{n}\bar{x} < -1.96 \cup \sqrt{n}\bar{x} > 1.96) = 0.05$$

$$I_{\sqrt{n}\bar{x} < -1.96 \cup \sqrt{n}\bar{x} > 1.96} \sim Bernoulli$$

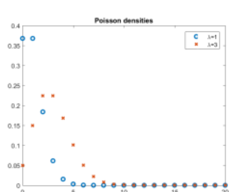
Suppose we draw n independent random variables $x_i \sim N(0,1)$ and generate the average $\bar{x}$. We know that $\Pr\left(\left|\bar{x}/\sqrt{1/n}\right| > 1.96\right) = 0.05$. Suppose we were to generate m such samples and count up the number of times that $\left|\bar{x}/\sqrt{1/n}\right| > 1.96$. Call this count divided by m, α.

a) What is the standard deviation of α?
b) Suppose we wanted there to be a 90 percent chance that the estimated α is between 0.04 and 0.06. How large must m be? You may approximate the distribution of α by the normal distribution.
c) Write a simulation program that sets $n = 10$ and uses a variety of values of m, some smaller than your value given in part (b) and some larger. Then run this for a number of times saving α. Make a plot that shows the theoretical standard deviation of α (from part (a))as a function of m as well as the empirical standard deviation of α from the simulations. Effectively, do a Monte Carlo of doing a Monte Carlo.

Poisson

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,\dots$$
$$\mathrm{E}(x) = \lambda$$
$$\mathrm{var}(x) = \lambda$$

$$f(x) = \frac{\lambda}{x} f(x-1),$$
$$x = 1, 2, \dots$$



Zero-inflated Poisson

$$f(x = 0) = \pi + (1-\pi)e^{-\lambda}$$
$$f(x|x > 0) = (1-\pi)\frac{e^{-\lambda}\lambda^x}{x!}$$

$$\mathrm{E}(x) = (1-\pi)\lambda$$
$$\mathrm{var}(x) = \lambda(1-\pi)(1+\lambda\pi)$$

Negative binomial

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, \dots$$

$$\mathrm{E}(x) = \frac{r}{p}$$

$$\mathrm{var}(x) = r\frac{1-p}{p^2}$$

n.b. The Poisson is a special case of the negative binomial with

$$\lambda = \frac{r}{1-p}, r \to \infty, p \to 1$$

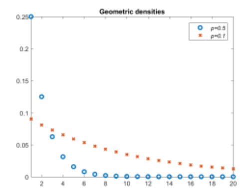
Geometric density

$$f(x) = p(1-p)^{x-1},$$
$$x = 1,2, \dots$$

$$\mathrm{E}(x) = \frac{1}{p}$$

$$\mathrm{var}(x) = \frac{1-p}{p^2}$$

where p is the probability of success in one trial



Geometric is memoryless

- Note that the probability of no success in n periods is $(1-p)^n$

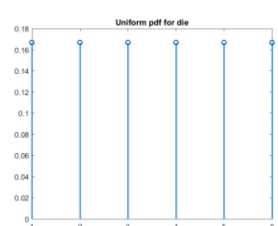
For $s > t$

$$f(x > s | x > t) = \frac{P(x > s \cap x > t)}{P(x > t)} = \frac{P(x > s)}{P(x > t)}$$

$$f(x > s | x > t) = \frac{(1-p)^s}{(1-p)^t} = (1-p)^{s-t}$$
$$= f(x > s - t)$$

Discrete uniform

$$f(x) = \frac{1}{n}, x = 1,2, \dots, n$$



Continuous Uniform

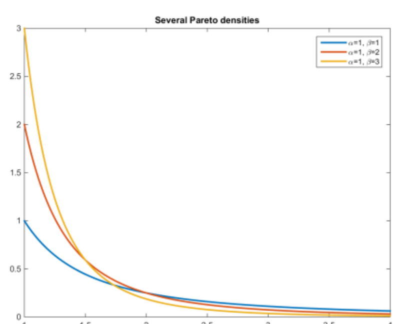
$$f(x|a,b) = \begin{cases} \dfrac{1}{b-a}, & \text{if } x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{E}(x) = \frac{b+a}{2}$$
$$\mathrm{var}(a) = \frac{(b-a)^2}{12}$$

Pareto distribution

$$f(x|\alpha,\beta) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \alpha < x < \infty, \alpha > 0, \beta > 0$$

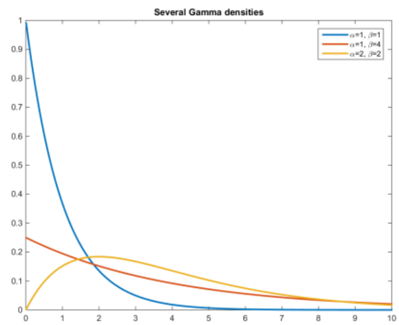
$$\mathrm{E}(x) = \frac{\beta\alpha}{\beta-1}, \beta > 1$$
$$\mathrm{var}(x) = \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)}, \beta > 2$$



Gamma

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, 0 < x < \infty, \alpha > 0, \beta > 0$$
$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$
$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$
$$\Gamma(n) = (n-1)! \text{ for integer } n$$
$$\mathrm{E}(x) = \alpha\beta$$
$$\mathrm{var}(x) = \alpha\beta^2$$

α is called the shape parameter. β is called the scale parameter.



χ_p^2 (chi-squared)

- χ_p^2 is gamma with $\beta = 2$ (and $\mathrm{p} = 2\alpha$).

$$f(x|p) = \frac{1}{\Gamma\left(\frac{p}{2}\right) 2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}, 0 < x$$

$$x \sim \chi_p^2$$
$$\mathrm{E}(x) = p$$
$$\mathrm{var}(x) = 2p$$

Write a Matlab program that draws a simulated $\chi^2(4)$ density and an analytic density on the same figure. To simulate the χ^2 generate 4 standard normals (randn(4,1)), square them and add them up. Do this a lot of times and then draw the histogram using the Matlab function histogram. Then draw the analytic expression using the chi2pdf function with 4 degrees of freedom.

Exponential

$$f(x) = \lambda e^{-\lambda x}, 0 < x < \infty$$
$$\mathrm{E}(x) = 1/\lambda$$
$$\mathrm{var}(x) = 1/\lambda^2$$

$$F(x) = 1 - e^{-\lambda x}$$

Exponential and discounting

- Suppose we are going to receive x dollars in t years discounted at rate r, then the present value would be
$$PV = xe^{-rt}$$
- Payment date is uncertain with exponential distribution

Exponential and discounting cont.

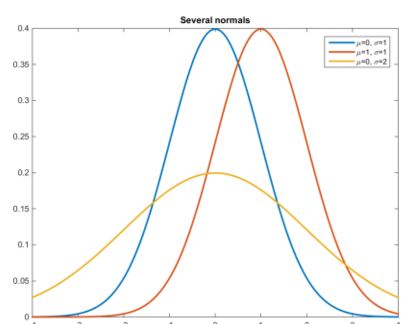
$$EPV = \int_0^\infty xe^{-rt} f(t)dt$$
$$EPV = \int_0^\infty xe^{-rt} \lambda e^{-\lambda t} dt$$
$$= x\frac{\lambda}{\lambda + r}\int_0^\infty (\lambda + r)e^{-(\lambda+r)t} dt$$
$$= x\frac{\lambda}{\lambda + r} \times 1 = \frac{x}{1 + r/\lambda}$$

Gaussian or Normal

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$
$$\mathrm{E}(x) = \mu$$
$$\mathrm{var}(x) = \sigma^2$$

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

The normal is symmetric around μ and has a maximum at μ.

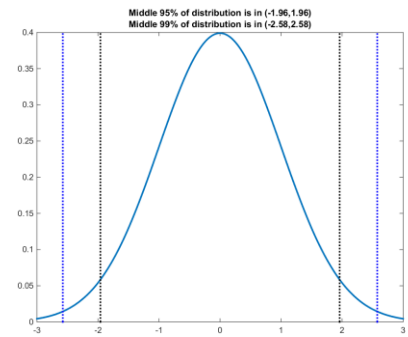


Standard normal

$$\phi(x) \equiv f_N(x|0,1)$$
$$\Phi(x) \equiv F_N(x|0,1)$$

- It is useful that

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$$
$$F(x|\mu,\sigma^2) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

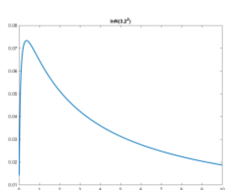


Log normal

If $\log x \sim N(\mu,\sigma^2)$, then $x \sim lnN(\mu,\sigma^2)$ is distributed $lognormal.$

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\log x - \mu)}$$

$$\mathrm{E}(x) = e^{\mu+\frac{\sigma^2}{2}}$$
$$\mathrm{var}(x) = (e^{\sigma^2}-1)e^{(2\mu+\sigma^2)}$$



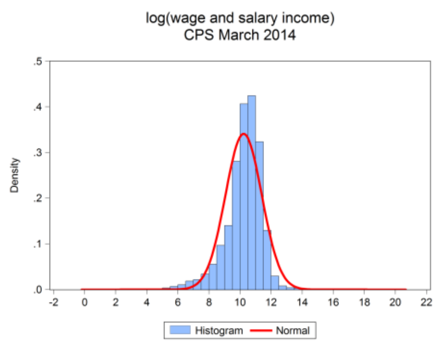
Example of proportional errors

- Cobb-Douglas with errors

$$Y = AK^{\beta}L^{1-\beta}\varepsilon$$

- If $\varepsilon \sim lnN(\cdot)$, then

$$\log Y = \log A + \beta \log K + (1-\beta)\log L + \log \varepsilon$$



Location and scale families

Theorem 3.5.1 Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants, then the function

$$g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf and

$$G(x|\mu,\sigma) = F\left(\frac{x-\mu}{\sigma}\right)$$

is a cdf.

Exponential families

- A pdf is an *exponential family* if it can be expressed as

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x)\right)$$

Includes: normal, gamma, beta, binomial, and Poisson

Logistic distribution

$$f(x) = \frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma \times \left(1+e^{-\frac{x-\mu}{\sigma}}\right)^2}$$

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

$$\mathrm{var}(x) = \frac{\sigma^2\pi^2}{3}$$

$$F(x) = \frac{1}{1+e^{-\frac{x-\mu}{\sigma}}}$$

Standard logistic distribution

$$F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{\sigma}}}$$

Standard logistic, $\mu = 0, \sigma = 1$ which gives

$$F(x) = \frac{1}{1 + e^{-x}}$$

An application of the standard logistic is if we think something happens if $x < x_0$ and otherwise doesn't happen. Then the *odds* of it happening are

$$\frac{F(x_0)}{1 - F(x_0)} = e^{x_0}, \log \frac{F(x_0)}{1 - F(x_0)} = x_0$$

Logistic regression on UW law school admissions

Dependent Variable: ADMIT
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 06/16/15 Time: 09:32
Sample: 1 1643 IF LSAT>99 AND LSAT<200 AND GPA<=4
Included observations: 1567
Convergence achieved after 4 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficien...	Std. Error	z-Statistic	Prob.
C	-10.02057	0.771661	-12.98571	0.0000
GPA	2.637745	0.218239	12.08650	0.0000

McFadden R-square...	0.100332	Mean dependent var	0.296107
S.D. dependent var	0.456685	S.E. of regression	0.428441
Akaike info criterion	1.095703	Sum squared resid	287.2737
Schwarz criterion	1.102540	Log likelihood	-856.4831
Hannan-Quinn criter.	1.098244	Deviance	1712.966
Restr. deviance	1903.998	Restr. log likelihood	-951.9992
LR statistic	191.0322	Avg. log likelihood	-0.546575
Prob(LR statistic)	0.000000		
Obs with Dep=0	1103	Total obs	1567
Obs with Dep=1	464		

Extreme value (Gumbel) Distribution

$$f(x) = \frac{1}{\beta} e^{-\left(\left(\frac{x-\mu}{\beta}\right)+e^{-\left(\frac{x-\mu}{\beta}\right)}\right)}$$

$$F(x) = e^{-e^{-\left(\frac{x-\mu}{\beta}\right)}}$$

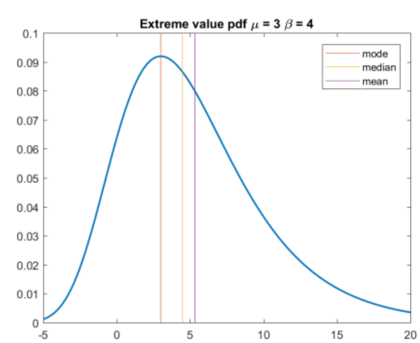
$$mode = \mu$$

$$median = \mu - \beta \log(\log(2))$$

$$mean = \mu + \beta\gamma$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.

$$\text{var} = \frac{\pi^2}{6}\beta^2$$



The pdf of the extreme value distribution can be written $f(x) = \frac{1}{\beta} e^{-\left(\left(\frac{x-\mu}{\beta}\right) + e^{-\left(\frac{x-\mu}{\beta}\right)}\right)}$, where $mean(x) = \mu + \beta\gamma, \gamma \approx 0.5772$ and $median(x) = \mu - \log(\log(2))$. Generate a bunch of samples of standard normals each with 120 observations. Find the maximum value for each sample. Find the mean and median of the distribution of the maxima. Now solve for the values of μ and β implied by the mean and median. Plot the empirical density (use the histogram function) of the maxima along with the theoretical pdf implied by the values of μ and β.