Serik Sagitov, Chalmers Tekniska Högskola, January 22, 2006

3. Continuous random variables

3.1 Probability density function

Def 1: continuous random variable

$X : \Omega \to (-\infty, \infty)$

is a real number resulting from a random experiment

Cdf

$F(x) = \mathrm{P}(X \le x)$ is a continuous function

increasing from 0 to 1 as x chages from $-\infty$ to $+\infty$

$\mathrm{P}(X = x) = 0$

Def 2: probability density function

pdf $f(x) = F'(x)$ describes a probability distribution

by a curve instead of a bar plot

Integrals instead of sums

$F(x) = \int_{-\infty}^{x} f(y)dy$, $\mathrm{P}(a < X < b) = \int_a^b f(y)dy$

$\mu = \int_{-\infty}^{\infty} xf(x)dx$

$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$

$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

3.2 Continuous uniform distribution

Uniform distribution with parameters a and b

$X \sim \mathrm{U}(a, b)$, $-\infty < a < b < \infty$

$f(x) = \frac{1}{b-a}$ for $a < x < b$ and $f(x) = 0$ otherwise

$\mu = \frac{a+b}{2}$, $\sigma^2 = \frac{(b-a)^2}{12}$, $\sigma = (b-a) \cdot 0.289$

Ex 1: systematic search

Number of trials required to open the lock $X \sim \mathrm{U}(10000)$

search time $T = \frac{X}{1000}$ h

if you do 1000 combinations per hour

Continuous approximate distribution

$T \sim \mathrm{U}(0, 10)$, $\mu = 5$ h, $\sigma =$ 2h 53 min

Mean deviation is smaller than the standard deviation σ

$\frac{1}{10} \int_0^{10} |x - 5| dx = \frac{1}{5} \int_0^5 (5 - x) dx =$ 2h 30 min

Compare $\mathrm{P}(T > 3) = 0.7$ and

$\mathrm{P}(T > 5 | T > 2) = \frac{\mathrm{P}(T>5)}{\mathrm{P}(T>2)} = \frac{(10-5)/10}{(10-2)/10} = 0.625$

3.3 Exponential distribution

Exponential distribution with parameter λ

$X \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$

$f(x) = \lambda e^{-\lambda x}$ for $x > 0$

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

A scale parameter λ, same shape for all λ

$\mathrm{Exp}(\lambda) = \frac{1}{\lambda} \mathrm{Exp}(1)$

Exponential approximation of the geometric distribution

if success is rare: p is small and n is large so that $np = \lambda$

then $\frac{1}{n}\mathrm{G}(p) \approx \mathrm{Exp}(\lambda)$

Interarrival times in the Poisson process

are independent r.v with the distribution $\mathrm{Exp}(\lambda)$

Lack of memory

$\mathrm{P}(X{>}a{+}b | X{>}b) = \frac{\mathrm{P}(X>a+b)}{\mathrm{P}(X>b)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda b}} = \mathrm{P}(X > a)$

Def 3: median value

Median M of a c.r.v X is another (instead of the mean) measure of central tendency defined by

$P(X > M) = P(X < M) = 0.5$

If distribution is symmetric, then median = mean

$$\boxed{\text{If } X \sim \text{Exp}(\lambda), \text{ then } M = \frac{\ln 2}{\lambda} = 0.693 \cdot \mu}$$

Ex 2: carbon-14 decay

$M = 5730$ years half-life of carbon-14

$\mu = 8267$ years

Ex 3: random search

Try the door codes at random

number of trials required $X \sim G(10^{-4})$

$P(X > 10000) = (0.9999)^{10000} = 0.37 \approx e^{-1}$

Search time

$T = \frac{X}{1000}$ hours, $T \sim \text{Exp}(0.1)$, $\mu = \sigma = 10$ hours

Continuous memoryless distribution

$P(T > 5 | T > 2) = \frac{e^{-0.5}}{e^{-0.2}} = e^{-0.3} = 0.741 = P(T > 3)$

3.4 Gamma distribution

Gamma distribution with parameters α and λ

$X \sim \text{Gamma}(\alpha, \lambda)$, shape $\alpha > 0$, scale $\lambda > 0$

$f(x) = \frac{1}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}\lambda e^{-\lambda x}$ for $x > 0$

$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}du$, $\Gamma(k) = (k-1)!$

$\mu = \alpha/\lambda$, $\sigma^2 = \alpha/\lambda^2$

Gamma and exponential distributions

$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda)$

if $X \sim \text{Gamma}(k, \lambda)$, then $X = Y_1 + \ldots + Y_k$

with independent $Y_i \sim \text{Exp}(\lambda)$

3.5 Normal distribution

Normal distribution with parameters μ and σ^2

$X \sim \text{N}(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ for $-\infty < x < \infty$

mean μ, variance σ^2

All normal distribution curves have the same shape

location parameter μ, scale parameter σ

Normal pdf

Symmetric "bell curve" centered at μ

graphical interpretation of σ: inflection points $\mu \pm \sigma$

$(\mu - 3\sigma) \underline{\ 2\%\ } (\mu - 2\sigma) \underline{14\%} (\mu - \sigma) \underline{34\%} (\mu)$

$(\mu) \underline{34\%} (\mu + \sigma) \underline{14\%} (\mu + 2\sigma) \underline{\ 2\%\ } (\mu + 3\sigma)$

Standardized random variable $\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

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

Standard normal distribution $Z \sim \mathrm{N}(0,1)$ zero mean $\mu = 0$, unit spread $\sigma = 1$

Cdf for $Z \sim \mathrm{N}(0,1)$

$\Phi(z) = \mathrm{P}(Z < z) = \int_{-\infty}^{x} \phi(y) dy$,

$\mathrm{P}(Z > z) = 1 - \Phi(z)$

$\mathrm{P}(Z < -z) = 1 - \Phi(z)$, $\mathrm{P}(|Z| > z) = 2(1 - \Phi(z))$

$\mathrm{P}(-z < Z < z) = 2\Phi(z) - 1$

Normal distribution table

z	1.00	1.28	1.64	1.96	2.00	2.33	2.58	3.00
$\Phi(z)$	.84	.90	.95	.975	.977	.99	.995	.9987
$2\Phi(z)-1$	.68	.80	.90	.95	.954	.98	.99	.9974

Three-sigma rule

99.74% of the $\mathrm{N}(\mu, \sigma^2)$ values are within $\mu \pm 3\sigma$

it requires on average $\frac{1}{(1-0.9974)} = 385$ observations

to see a three-sigma outlier

A useful approximation for large z

$1 - \Phi(z) \approx \frac{1}{z\sqrt{2\pi}} e^{-z^2/2}$

Ex 4: Intellegence Quotient

Given IQ $\sim$ N(100,15^2) find

P(IQ < 85), P(IQ > 115), P(IQ > 130), P(IQ > 145)

P(|IQ – 100| > 45), P(IQ > 175) $= 3 \cdot 10^{-7}$

3.6 Transformed r.v.

Given pdf $f(x)$ for X and non-decreasing function $h(x)$

compute pdf $g(x)$ for the transformed r.v. $Y = h(X)$

$G(y) = \mathrm{P}(Y \leq y) = F(h_{-1}(y))$

$g(y) = \frac{d}{dy}F(h_{-1}(y)) = \frac{f(h_{-1}(y))}{h'(h_{-1}(y))}$

if non-increasing function $h(x)$, then $g(y) = \frac{f(h_{-1}(y))}{|h'(h_{-1}(y))|}$

Ex 5: linear transformation

Linear transformation $Y = aX + b$

$h_{-1}(y) = \frac{y-b}{a}$, $g(y) = \frac{1}{|a|}f(\frac{y-b}{a})$

Ex 6: squared $\mathbf{N}(0,1)$

For $Y = Z^2$ with $Z \in \mathrm{N}(0,1)$

$h(y) = y^2$ is not monotone

$G(y) = \mathrm{P}(Y \leq y) = 2\mathrm{P}(0 < Z \leq \sqrt{y}) = 2\Phi(\sqrt{y}) - 1$

$g(y) = \frac{1}{\sqrt{y}}\phi(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}}e^{-y/2}$ Gamma($\frac{1}{2}, \frac{1}{2}$) pdf

Ex 7: infinite mean

For $Y = 1/U$ with $U \in \mathrm{U}(0,1)$

$h_{-1}(y) = 1/y$, $g(y) = \frac{1}{y^2}$ for $y > 1$

$\mathrm{E}(Y) = \int_1^\infty dy/y = \infty$

Generating a distribution

Given a cdf $F(x)$ simulate $X \sim F(x)$ with $X = F_{-1}(U)$

using a pseudorandom number $U \in \mathrm{U}(0,1)$

Ex 8: generating exponential distribution

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

$X = -\ln(1-U)/\lambda$ or even $X = -\ln(U)/\lambda$

3.7 Mean and variance of a transformed r.v.

Compute the mean and variance of $Y = h(X)$

in terms of $\mu = \mathrm{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$

If h is linear $Y = aX + b$

then $\mathrm{E}(Y) = a\mu + b$, $\mathrm{Var}(X) = a^2\sigma^2$

If h is quadratic $Y = aX + b + c(X - \mu)^2$

then $\mathrm{E}(Y) = a\mu + b + c\sigma^2$

Propagation of error

Approximate method when h is linear or quadratic

in a high probability range of X

The first order Taylor series expansion

$h(x) \approx h(\mu) + (x - \mu)h'(\mu)$

$\mathrm{E}(Y) \approx h(\mu)$, $\mathrm{Var}(X) \approx (h'(\mu))^2\sigma^2$

The second order Taylor series expansion

$h(x) \approx h(\mu) + (x - \mu)h'(\mu) + \frac{1}{2}(x - \mu)^2 h''(\mu)$

$\mathrm{E}(Y) \approx h(\mu) + \frac{1}{2}h''(\mu)\sigma^2$

3.8 Probability distribution quiz

Suggest a probability distribution or pmf/pdf shape for

1. Number of children untill the first son
2. Waiting time for a bus
3. Your daily expenses
4. The number of matches when guessing 6 out of 49
5. The next digit in the number $\pi = 3.1415926535897$
6. 10 people throw out fingers, total number of fingers
7. Human lifelength