Serik Sagitov, Chalmers Tekniska Högskola, May 13, 2005

4. Joint distributions

4.1 Discrete joint distributions

Def 1: joint and marginal pmf

For a pair of random counts X and Y define

the joint pmf $p_{XY}(n,k) = \mathrm{P}(X = n, Y = k)$

and two marginal pmf

$p_X(n) = \mathrm{P}(X = n)$ and $p_Y(k) = \mathrm{P}(Y = k)$

Computing marginal pmf from the joint pmf

$p_X(n) = \sum_k p_{XY}(n,k)$, $p_Y(k) = \sum_n p_{XY}(n,k)$

Def 2: conditional pmf

For a pair of random counts X and Y define

conditional pmf of Y given X: $p_{Y|X}(k|n) = \frac{p_{XY}(n,k)}{p_X(n)}$

conditional pmf of X given Y: $p_{X|Y}(n|k) = \frac{p_{XY}(n,k)}{p_Y(k)}$

Def 3: independent random counts

X and Y are independent if $p_{XY}(n,k) = p_X(n)p_Y(k)$

for all possible combinations (n,k)

Ex 1: die-coin experiment

step one: roll a fair die, $X =$ the die score

step two: toss a fair coin X times, $Y = \#\{\text{heads}\}$

The joint distribution

$p_X(n) = \frac{1}{6}$, $n = 1, 2, 3, 4, 5, 6$

$p_{Y|X}(k|n) = \binom{n}{k} 2^{-n}$, $k = 0, \ldots, n$

by the LTP $p_{XY}(n,k) = \frac{1}{6}\binom{n}{k} 2^{-n}$

$p_{XY}(n,k)$	k=0	k=1	k=2	k=3	k=4	k=5	k=6	$p_X(n)$
n=1	.083	.083	0	0	0	0	0	.167
n=2	.042	.083	.042	0	0	0	0	.167
n=3	.021	.063	.063	.021	0	0	0	.167
n=4	.010	.042	.063	.042	.010	0	0	.167
n=5	.005	.026	.052	.052	.026	.005	0	.167
n=6	.003	.016	.039	.052	.039	.016	.003	.167
$p_Y(k)$	.164	.313	.258	.167	.076	.021	.003	1

Compare the prior distribution $X \sim \text{dU}(6)$
with the posterior distribution of X given Y

$p_{X\|Y}(n\|k)$	k=0	k=1	k=2	k=3	k=4	k=5	k=6
n=1	.508	.265	0	0	0	0	0
n=2	.254	.265	.162	0	0	0	0
n=3	.127	.201	.243	.126	0	0	0
n=4	.064	.134	.243	.252	.133	0	0
n=5	.032	.083	.201	.311	.347	.238	0
n=6	.016	.051	.151	.311	.520	.762	1
total	1	1	1	1	1	1	1

4.2 Continuous joint distributions

Def 4: joint pdf

Joint pdf $f_{XY}(x,y)$ for two c.r.v. X and Y is defined by
$\text{P}(a < X < b, c < Y < d) = \int_a^b \int_c^d f_{XY}(x,y)dxdy$
can be viewed as a surface or a scatter plot
a top of the surface = a cluster in the scatter plot

The marginal density of X: $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$

Conditional pdf $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$
profile of the pdf surface cut by vertical plane $\{X = x\}$

Def 5: independent random variables
X and Y are independent if $f_{XY}(x, y) = f_X(x) f_Y(y)$

Ex 2: uniform distribution over the unit square
Random point (X, Y) on a plain takes values (x, y) in
$S = \{0 < x < 1,\ 0 < y < 1\}$ with $f_{XY}(x, y) = 1$
Marginal distributions: $X \sim$ U(0,1), $Y \sim$ U(0,1)
X and Y are independent since
$f_{XY}(x, y) = 1$, and $f_X(x) = 1$, $f_Y(y) = 1$
Ex 3: uniform distribution over a disk
Random point (X, Y) on a plain takes values (x, y) in
$D = \{x^2 + y^2 < 1\}$ with $f_{XY}(x, y) = \frac{1}{\pi} = 0.318$
Marginal pdf
$f_X(x) = \frac{2}{\pi}\sqrt{1 - x^2}, \quad f_Y(y) = \frac{2}{\pi}\sqrt{1 - y^2}$
Negative dependence between $|X|$ and $|Y|$:
when $|X|$ is closer to 1, $|Y|$ gets closer to 0
$f_{XY}(x, y) \neq f_X(x) f_Y(y)$

4.3 Conditional expectation

Conditional expectation and variance of X given Y are random variables $\mathrm{E}(X|Y)$ and $\mathrm{Var}(X|Y)$

Ex 1: die-coin experiment

Y values	0	1	2	3	4	5	6
$\mathrm{E}(X\|Y)$ values	1.90	2.65	3.94	4.81	5.38	5.75	6.00
$\mathrm{Var}(X\|Y)$ values	1.44	2.14	1.64	1.03	0.87	0.18	0.00
probabilities	.164	.313	.258	.167	.076	.021	.003

$\mathrm{E}(\mathrm{E}(X|Y) = 1.90 \cdot 0.164 + 2.65 \cdot 0.313 + 3.94 \cdot 0.258+$
$4.81 \cdot 0.167 + 5.38 \cdot 0.076 + 5.75 \cdot 0.021 + 6.00 \cdot 0.003 = 3.5$
$\mathrm{Var}(\mathrm{E}(X|Y)) = 1.35$, $\mathrm{E}(\mathrm{Var}(X|Y)) = 1.57$

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

$\mathrm{Var}(\mathrm{E}(X|Y)) + \mathrm{E}(\mathrm{Var}(X|Y))$=2.92= $\mathrm{Var}(X)$

Laws of Total Expectation and Total Variance

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

Proof of the law of total variance

$\mathrm{Var}(Y) = \mathrm{E}(\mathrm{E}((Y - \mu_Y)^2|X))$
$= \mathrm{E}(\mathrm{E}((Y - \mu_{Y|X})^2|X)) + \mathrm{E}(\mathrm{E}((\mu_{Y|X} - \mu_Y)^2|X))$
$= \mathrm{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mu_{Y|X})$

Ex 4: systematic vs random search

Two independent r.v. $S \in \mathrm{U}(0, 10)$ and $R \in \mathrm{Exp}(0.1)$

$\mathrm{P}(R > S) = \mathrm{P}(R > S) = \mathrm{E}(\mathrm{P}(R > S|S))$
$= \mathrm{E}(e^{-0.1S}) = 0.1 \int_0^{10} e^{-0.1x}dx = \int_0^1 e^{-x}dx = 0.632$

Optimal predictor

Optimal predictor of Y based on X is $\hat{Y} = \mathrm{E}(Y|X)$

minimal mean square error $\mathrm{E}(Y - \hat{Y})^2 = \mathrm{E}(\mathrm{Var}(Y|X))$

non-linear trend of Y for a certain range of X values

Two sources of variation in Y

$\mathrm{Var}(\mathrm{E}(Y|X))$ variation caused by X

$\mathrm{Var}(Y|X) = \mathrm{E}((Y - \hat{Y})^2|X)$ scatter around the trend

4.4 Covariance and correlation

Two measures of association between X and Y

covariance $\mathrm{Cov}(X, Y)$ and correlation coefficient ρ

$$\boxed{\mathrm{Cov}(X, Y) = \mathrm{E}(X - \mu_X)(Y - \mu_Y)}$$

Addition rule of variance

$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$

$$\boxed{\text{Correlation coefficient: } \rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y},\ -1 \le \rho \le 1}$$

Compute covariance by $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mu_X \mu_Y$

ρ is a scale free degree of straight-line association

X and Y are called uncorrelated if $\rho = 0$

$$\boxed{\begin{array}{r}\text{If } X \text{ and } Y \text{ are uncorrelated, then } \mathrm{E}(XY) = \mu_X \mu_Y \\ \text{and } \mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)\end{array}}$$

Ex 3: uniform distribution over a disk

Law of total expectation

$\mathrm{E}(Y|X) = 0$, $\mathrm{E}(X \cdot Y) = \mathrm{E}(X \cdot \mathrm{E}(Y|X)) = 0$

$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mu_X \mu_Y = 0$, $\rho = 0$

Independent r.v. are uncorrelated, but the converse is not always true

Optimal linear predictor

Optimal linear predictor of Y given X is

$\tilde{Y} = \mu_Y + \frac{\sigma_Y}{\sigma_X} \cdot \rho \cdot (X - \mu_X)$

Ex 1: die-coin experiment

k	0	1	2	3	4	5	6	8
$P(XY = k)$	.164	.083	.083	.063	.083	.026	.079	.063

9	10	12	15	16	18	20	24	30	36
.021	.052	.081	.052	.010	.052	.026	.039	.016	.003

$E(Y) = 1.75$, $\text{Var}(Y) = 1.61$, $E(XY) = 7.6$

$\text{Cov}(X, Y) = 7.6 - 3.5 \cdot 1.75 = 1.48$, $\rho = \frac{1.46}{1.71 \cdot 1.27} = 0.68$

$\text{Var}(X + Y) = 2.92 + 1.61 + 2 \cdot 1.48 = 7.49$

$\tilde{X} = 3.5 + 0.91 \cdot (Y - 1.75) = 1.91 + 0.91 \cdot Y$

Y	0	1	2	3	4	5	6
$\hat{X}$	1.90	2.65	3.94	4.81	5.38	5.75	6.00
$\tilde{X}$	1.91	2.82	3.73	4.64	5.55	6.46	7.37

Ex 5: height prediction

Mother, Father, Daughter, Son:

$\tilde{H}_S = \frac{1}{2}(H_M + 13\text{cm} + H_F)$

$\tilde{H}_D = \frac{1}{2}(H_M - 13\text{cm} + H_F)$

4.5 Multinomial distribution

Multinomial distribution with parameters $n, p_1, \ldots, p_r$

$(X_1, \ldots, X_r) \sim \mathrm{Mn}(n; p_1, \ldots, p_r)$

$n \in \mathbb{N}$, $p_1 + \ldots + p_r = 1$

$\mathrm{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} \ldots, p_r^{k_r}$

where $k_1 + \ldots + k_r = n$

Def 6: multinomial trials

independent repeated experiments with r possible outcomes with probabilities $p_1, \ldots, p_r$

X_i = number of outcomes of type i in n trials

Marginal distributions

$X_i \sim \mathrm{Bin}(n; p_i)$, $\mathrm{E}(X_i) = np_i$, $\mathrm{Var}(X_i) = np_iq_i$

Two cells combined $X_i + X_j \sim \mathrm{Bin}(n; p_i + p_j)$

$\mathrm{Var}(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$

Addition rule rule of variance

$\mathrm{Cov}(X_i, X_j) = \frac{\mathrm{Var}(X_i+X_j) - \mathrm{Var}(X_i) - \mathrm{Var}(X_j)}{2} = -np_ip_j$

X_i and X_j are negatively correlated $\rho_{ij} = -\sqrt{\frac{p_ip_j}{q_iq_j}}$

Ex 6: YATZY game

Roll five fair dice: six random variables

X_i = number of "i"s, $i = 1, 2, 3, 4, 5, 6$

$(X_1, \ldots, X_6) \in \mathrm{Mn}(5; \frac{1}{6}, \ldots, \frac{1}{6})$

Negative correlation between X_i and X_j

$\rho = -\sqrt{\frac{1/6}{5/6}}\sqrt{\frac{1/6}{5/6}} = -0.2$

Probabilities of different combinations

$$\mathrm{P}(5) = 6 \cdot \mathrm{P}(X_1 = 5) = 6 \cdot (\tfrac{1}{6})^5 = 0.0008$$
$$\mathrm{P}(4{+}1) = 6 \cdot \mathrm{P}(X_1 = 4) = 6 \cdot \tbinom{5}{4}(\tfrac{1}{6})^4(\tfrac{5}{6}) = 0.0193$$
$$\mathrm{P}(3+2) = 6 \cdot 5 \cdot \mathrm{P}(X_1 = 3, X_2 = 2, X_3 = \ldots = 0)$$
$$= 30 \cdot \tbinom{5}{3}(\tfrac{1}{6})^5 = 0.0386$$
$$\mathrm{P}(3+1+1) = 6 \cdot \tbinom{5}{2} \cdot \mathrm{P}(X_1 = 3, X_2 = 1, X_3 = 1)$$
$$= 60 \cdot \tbinom{5}{3,1,1}(\tfrac{1}{6})^5 = 0.1543$$
$$\mathrm{P}(2+2+1) = 6 \cdot \tbinom{5}{2} \cdot \mathrm{P}(X_1 = 2, X_2 = 2, X_3 = 1)$$
$$= 60 \cdot \tbinom{5}{2,2,1}(\tfrac{1}{6})^5 = 0.2315$$
$$\mathrm{P}(2{+}1{+}1{+}1) = 6 \cdot \tbinom{5}{3} \cdot \mathrm{P}(X_1 = 2, X_2 = X_3 = X_4 = 1)$$
$$= 60 \cdot \tbinom{5}{2,1,1,1}(\tfrac{1}{6})^5 = 0.4630$$
$$\mathrm{P}(1{+}1{+}1{+}1{+}1) = 6 \cdot \mathrm{P}(X_1 = \ldots = X_5 = 1) = 0.0924$$

Ex 7: students' ID numbers

$X_i = \#\{\text{students whose ID number ends with } i\}$
find the joint distribution of $(X_0, \ldots, X_9)$
test in the class

4.6 Bivariate normal distribution

Bivariate normal distribution

$$(X, Y) \sim \mathrm{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\}$$

Five parameters

location parameters μ_1, μ_2 any values

scale parameters σ_1, σ_2 positive

shape parameter $-1 < \rho < 1$ = correlation coefficient

Marginal distributions: $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$

With the normal joint distribution OP = OLP
$\mathrm{E}(Y\|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$

Draw ellipse contour plots for $\rho > 0$, $\rho = 0$, $\rho < 0$

$\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} = \text{constant}$

together with regression lines

Coefficient of determination $\rho^2 = \frac{\mathrm{Var}(E(Y|X))}{\mathrm{Var}(Y)}$

proportion of variation in Y

explained by variation in X

Conditonal distribution

$(Y|X) \sim \mathrm{N}(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1), \sigma_2^2(1 - \rho^2))$

Ex 8: heights of fathers and sons

http://www.scc.ms.unimelb.edu.au/discday/dyk/faso.html

data: $n = 1078$ paired observations of two variables

X = father's height in inches

Y = son's height in inches

Parameters estimated from the data

$\mu_1 = 68$, $\sigma_1 = 2.7$ (1 inch = 2.54 cm)

$\mu_2 = 69$, $\sigma_2 = 2.7$, $\rho = 0.5$

4.7 Moment generating function

Def 7: mgf

Moment generating function of a r.v. X is the function

$M(t) = \mathrm{E}(e^{tX})$ if exists for $0 < t < \epsilon$

Properties of mgf

$M(t)$, $0 < t < \epsilon$ uniquely determines the cdf

$\mathrm{E}(X^r) = M^{(r)}(0)$, $r \in \mathbb{N}_0$

$\mathrm{E}(e^{t(X+Y)}) = \mathrm{E}(e^{tX})\mathrm{E}(e^{tY}) \Leftrightarrow X$ and Y are indep

Discrete distributions

Bernoulli(p): $M(t) = e^t p + e^0 q = q + pe^t$

Bin(n,p): $M(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} = (q+pe^t)^n$

Geom(p): $M(t) = \sum_{k=1}^{n} e^{tk} pq^{k-1} = \frac{pe^t}{1-qe^t}$, $t < |\ln(q)|$

Nb(r,p): $M(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$

Pois(λ): $M(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(e^t-1)}$

Continuous distributions

Exp(λ): $M(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}$, $t < \lambda$

Gamma(α,λ): $\int_0^\infty e^{tx} \lambda \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} dx = \left(\frac{\lambda}{\lambda-t}\right)^\alpha$

N(0,1): $M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{t^2/2}$

Randomized Poisson

suppose $X \sim$ Gamma(r,λ) and $(Y|X) \sim$ Pois(X)

$$\mathrm{E}(e^{t(Y+r)}) = e^{tr}\mathrm{E}(\mathrm{E}(e^{tY})|X) = e^{tr}\mathrm{E}(e^{X(e^t-1)})$$

$$= e^{tr}\left(\frac{\lambda}{\lambda-(e^t-1)}\right)^r = \left(\frac{\frac{\lambda}{1+\lambda}e^t}{1-\frac{1}{1+\lambda}e^t}\right)^r$$

thus $(Y+r) \sim$ Nb($r, \frac{\lambda}{1+\lambda}$)

4.8 Expectation and variance of a ratio

Let $Z = h(X,Y)$, where

$E(X) = \mu_1$, $E(Y) = \mu_2$, $\mathrm{Var}(X) = \sigma_1^2$, $\mathrm{Var}(Y) = \sigma_2^2$

$\mathrm{Cov}(X,Y) = \rho\sigma_1\sigma_2$

Function $h(x,y)$ around (μ_1, μ_2)

$h_0 = h(\mu_1,\mu_2)$, $h_x = h'_x(\mu_1,\mu_2)$, $h_y = h'_y(\mu_1,\mu_2)$

$h_{xx} = h''_{xx}(\mu_1,\mu_2)$, $h_{yy} = h''_{yy}(\mu_1,\mu_2)$, $h_{xy} = h''_{xy}$

Propagation of error

$$Z = h(X,Y) \approx h + (X-\mu_1)h_x + (Y-\mu_2)h_y$$

$$\mathrm{Var}(Z) \approx (\sigma_1 h_x)^2 + (\sigma_2 h_y)^2 + 2\rho\sigma_1\sigma_2 h_x h_y$$

$$Z \approx h + (X-\mu_1)h_x + (Y-\mu_2)h_y + \tfrac{1}{2}(X-\mu_1)^2 h_{xx}$$
$$+\tfrac{1}{2}(Y-\mu_2)^2 h_{yy} + (X-\mu_1)(Y-\mu_2)^2 h_{xy}$$

$$E(Z) \approx h + \tfrac{1}{2}\sigma_1^2 h_{xx} + \tfrac{1}{2}\sigma_2^2 h_{yy} + \rho\sigma_1\sigma_2 h_{xy}$$

The ratio $Z = h(X,Y)$, where $h(x,y) = y/x$

$h_0 = \mu_2/\mu_1$ $\quad h_x = -\mu_2/\mu_1^2$ $\quad h_y = 1/\mu_1$

$h_{xx} = 2\mu_2/\mu_1^3$ $\quad h_{yy} = 0$ $\quad h_{xy} = -1/\mu_1^2$

$$E(Y/X) \approx \tfrac{\mu_2}{\mu_1} + \sigma_1^2\tfrac{\mu_2}{\mu_1^3} - \rho\sigma_1\sigma_2\tfrac{1}{\mu_1^2}$$

$$\mathrm{Var}(Y/X) \approx \sigma_1^2\tfrac{\mu_2^2}{\mu_1^2} + \sigma_2^2\tfrac{1}{\mu_1^2} - 2\rho\sigma_1\sigma_2\tfrac{\mu_2}{\mu_1^3}$$

Discuss the effects

of small standard deviations σ_1 and σ_2

positive and negative correlation

small μ_1