Serik Sagitov, Chalmers Tekniska Högskola, August 16, 2004

3. Joint distributions

3.1 Discrete joint distributions

Def 1: joint pmf of X and Y

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

Marginal distributions:
$p_X(n) = \Sigma_k\, p_{XY}(n,k)$ and $p_Y(k) = \Sigma_n\, p_{XY}(n,k)$

Def 2: independent random variables

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

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

Ex 1: die-coin experiment

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

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

Conditional and joint distributions

$p_{Y|X}(k|n) = \binom{n}{k}2^{-n}$ and $p_{XY}(n,k) = \frac{1}{6}\binom{n}{k}2^{-n}$

$k = 0,\ldots,n$, where $n = 1,2,3,4,5,6$

$p_{XY}(n,k)$	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 $X \sim \mathrm{dU}(6)$
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

3.2 Optimal predictor

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

Laws of Total Expectation and Total Variance

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

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

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

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

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

$\mathrm{Var}(X) = 2.92 = 1.35 + 1.57$

$\mathrm{Var}(\mathrm{E}(X|Y)) = 1.35$, $\mathrm{E}(\mathrm{Var}(X|Y)) = 1.57$

3.3 Covariance and correlation

Two measures of association between X and Y

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

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 $\rho = 0$

$$\boxed{\begin{array}{l}\text{Optimal linear predictor of } X \text{ given } Y \text{ is} \\ \qquad \tilde{X} = \mu_X + \frac{\sigma_X}{\sigma_Y} \cdot \rho \cdot (Y - \mu_Y)\end{array}}$$

Ex 1: die-coin experiment

k	0	1	2	3	4	5	6	8
$\mathrm{P}(XY=k)$	.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

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

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

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

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

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 2: family size model

Number of children in a family $Z \sim \text{Pois}(2)$, $\text{E}(Z) = 2$

probabilities in percentages:

n	0	1	2	3	4	5
$p(n)$	13.5	27.1	27.1	18.0	9.02	3.61

n	6	7	8	9	10	11
$p(n)$	1.20	0.34	0.09	0.02	0.00	0.00

Modelling number of girls X and boys Y in the family:

given $Z = n$ toss a fair coin n times

$X = \#\{\text{heads}\}$, $Y = \#\{\text{tails}\}$

Conditional distributions

$X \sim \text{Bin}(n, 0.5)$, $Y \sim \text{Bin}(n, 0.5)$

Joint pmf of X and Y in percentages:

	0	1	2	3	4	5	6	$p_X(n)$
0	13.5	13.5	6.77	2.26	0.56	0.11	0.02	36.8
1	13.5	13.5	6.77	2.26	0.56	0.11	0.02	36.8
2	6.77	6.77	3.38	1.13	0.28	0.06	0.01	18.4
3	2.26	2.26	1.13	0.38	0.09	0.02	0.00	6.13
4	0.56	0.56	0.28	0.09	0.02	0.00	0.00	1.53
5	0.11	0.11	0.06	0.02	0.00	0.00	0.00	0.31
6	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.05
$p_Y(k)$	36.8	36.8	18.4	6.13	1.53	0.31	0.05	1

X and Y are independent

marginal distributions $X \sim \text{Pois}(1)$, $Y \sim \text{Pois}(1)$

3.4 Multinomial distribution

Def 3: 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

$$\boxed{\begin{array}{c}\text{Multinomial } (X_1, \ldots, X_r) \sim \text{Mn}(n; p_1, \ldots, p_r) \\ \text{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} \ldots, p_r^{k_r} \\ \text{where } p_1 + \ldots + p_r = 1,\ k_1 + \ldots + k_r = n\end{array}}$$

Marginal distributions

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

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

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

Addition rule rule of variance

$\text{Cov}(X_i, X_j) = \frac{\text{Var}(X_i + X_j) - \text{Var}(X_i) - \text{Var}(X_j)}{2} = -np_i p_j$

$$\boxed{X_i \text{ and } X_j \text{ are negatively correlated } \rho_{ij} = -\sqrt{\frac{p_i p_j}{q_i q_j}}}$$

Ex 3: genotype frequences

Population with genotypes AA = 1, Aa = 2, aa = 3

having frequencies $p_1 = 0.64$, $p_2 = 0.32$, $p_3 = 0.04$

sample $n = 8$ individuals with replacement

$\text{P}(1, 1, 1, 1, 1, 2, 2, 3) = p_1^5 p_2^2 p_3^1 = 0.00044$

Genotypic counts: $(X_1, X_2, X_3) \sim \text{Mn}(n;\ p_1,\ p_2,\ p_3)$

$\text{P}(X_1 = 5, X_2 = 2, X_3 = 1) = \binom{8}{5,2,1} p_1^5 p_2^2 p_3^1 = 0.074$

Ex 4: students' ID numbers

$X_i = \#\{\text{students whose ID number ends with } i\}$

find the joint distribution of $(X_0, \ldots, X_9)$

Ex 5: Wright-Fisher model

Constant population size N

X_i = offspring number of i-th individual

$(X_1, \ldots, X_N) \sim \text{Mn}(N; \frac{1}{N}, \ldots, \frac{1}{N})$

Negative correlation between X_i and X_j

$\rho = -\frac{1}{N-1}$ vanishes for large population

Common marginal distribution

$X_i \sim \text{Bin}(N; \frac{1}{N})$, $\mu = 1$, $\sigma^2 = \frac{N-1}{N}$

Poisson approximation for large N

$X_i \approx \text{Pois}(1)$, $\mu = \sigma^2 = 1$

3.5 Continuous joint distributions

Joint pdf of X and Y

$$f_{XY}(x,y) = \frac{\text{P}(x<X<x+dx,\ y<Y<y+dy)}{dx \cdot dy}$$

can be viewed as a surface or a scatter plot

a top of the surface = a cluster in the scatter plot

$$\text{P}((X,Y) \in A) = \int_A f_{XY}(x,y)dxdy$$

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

Ex 6: 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 7: 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
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

3.6 Bivariate normal distribution

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

location parameters μ_1, μ_2

scale parameters σ_1, σ_2

shape parameter $\rho =$ 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 $\rho > 0$, $\rho = 0$, $\rho < 0$

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)

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

$\rho = 0.5$