PROBABILITY FOR STATISTICS: EXAM SOLUTIONS 2014-15

Q1.(i) $P_{X+Y}(s) = \sum_{n=0}^{\infty} P(X+Y=n)s^n$.

$P(X+Y=n) = \sum_{k=0}^{n} P(X=k, Y=n-k) = \sum_{k=0}^{n} P(X=k)P(Y=n-k)$, by independence. Substitute, and put $j := n-k$ to get

$$P_{X+Y}(s) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} P(X=k).P(Y=j).s^j.s^k$$

$$= \sum_{k=0}^{\infty} P(X=k)s^k. \sum_{j=0}^{\infty} P(Y=j).s^j = P_X(s).P_Y(s). \qquad \mathbf{[5]}$$

(ii) $P_X(s) = \sum_{n=0}^{\infty} e^{-\lambda}(\lambda^n/n!).s^n = e^{-\lambda}\sum_{n=0}^{\infty}(\lambda s)^n/n!$
$= e^{-\lambda}.e^{\lambda s} = e^{-\lambda(1-s)}$. **[2]**

(iii) Combining, $P_{X+Y}(s) = e^{-\lambda(1-s)}.e^{-\mu(1-s)} = e^{-(\lambda+\mu)(1-s)}$. So $X+Y \sim P(\lambda+\mu)$. **[3]**

(iv) $EX = \sum_{n=0}^{\infty} n.P(X=n)$. One can evaluate the sum directly, but the easiest way to get the sum is to differentiate the generating function and evaluate it at $s=1$ (proof: one can take the d/ds inside the sum; $d(s^n)/ds = ns^{n-1}$; this gives the factor n on putting $s=1$). As $d[e^{-\lambda(1-s)}]/ds = \lambda e^{-\lambda(1-s)}$, this gives $EX = \lambda$: the mean of a Poisson random variable is its parameter (its variance is λ too, but we won't need this here). Then $E(X+Y) = EX + EY = \lambda + \mu$ follows by linearity of expectation E (expectation is integration, and integration is linear). **[6]**

(v) $X = (X_t)$ is a $Ppp(\lambda)$ if for any measurable set A (equivalently, for any interval A), the number $X(A)$ of points of the point process X in A is Poisson distributed with parameter $\lambda|A|$, $X(A) \sim P(\lambda|A|)$, and the numbers of points of X in disjoint sets are independent. **[3]**

(vi) If $X \sim Ppp(\lambda)$, $Y \sim Ppp(\mu)$, then for any A, $X(A) \sim P(\lambda|A|)$, $Y(A) \sim P(\mu|A|)$, and these are independent as X, Y are independent. So $(X+Y)(A) \sim P((\lambda+\mu)|A|)$, by (iii). Also, for disjoint sets A, B, $X(A)$, $X(B)$ are independent as X is Poisson, and similarly so are $Y(A)$, $Y(B)$, while both X-counts are independent of both Y-counts as X and Y are independent. Combining, $(X+Y)(A)$ and $(X+Y)(B)$ are independent. This completes the proof that $X+Y$ is $Ppp(\lambda+\mu)$. **[6]**

[Similar seen – Lectures and Problems]

Q2. (i) $r := S_{xy}/(S_x S_y)$, where $S_{xy} := \overline{(x-\overline{x})(y-\overline{y})}$, $S_x := \sqrt{S_{xx}}$ etc. **[2]**
$\rho := \sigma_{xy}/(\sigma_x\sigma_y)$, where $\sigma_{xy} := E[(x-Ex)(y-Ey)]$, $\sigma_x := \sqrt{\sigma_{xx}}$ etc. **[2]**
(ii) For r, consider the quadratic

$$\begin{aligned} Q(\lambda) &= \lambda^2\frac{1}{n}\sum_1^n (x_i-\bar{x})^2 + 2\lambda\frac{1}{n}\sum_1^n (x_i-\bar{x})(y_i-\bar{y}) + \frac{1}{n}\sum_1^n (y_i-\bar{y})^2 \\ &= \lambda^2\overline{(x-\bar{x})^2} + 2\lambda\overline{(x-\bar{x})(y-\bar{y})} + \overline{(y-\bar{y})^2} \\ &= \lambda^2 S_{xx} + 2\lambda S_{xy} + S_{yy}. \end{aligned}$$

Now $Q(\lambda) \geq 0$ for all λ, so $Q(\cdot)$ is a quadratic which does not change sign. So its discriminant is ≤ 0 (if it were > 0, there would be distinct real roots and a sign change). So ("$b^2 - 4ac \leq 0$")

$$s_{xy}^2 \leq s_{xx}s_{yy} = s_x^2 s_y^2, \qquad r^2 := (s_{xy}/s_x s_y)^2 \leq 1: \qquad -1 \leq r \leq +1. \qquad \textbf{[6]}$$

Similarly for ρ, from

$$\begin{aligned} Q(\lambda) &= E[\lambda^2(x-Ex)^2 + 2\lambda(x-Ex)(y-Ey) + (y-Ey)^2] \\ &= \lambda^2 E[(x-Ex)^2] + 2\lambda E[(x-Ex)(y-Ey)] + E[(y-Ey)^2] \\ &= \lambda^2\sigma_x^2 + 2\lambda\sigma_{xy} + \sigma_y^2. \end{aligned}$$

As before $Q(\lambda) \geq 0$ for all λ, as the discriminant is ≤ 0, i.e.

$$\sigma_{xy}^2 \leq \sigma_x^2\sigma_y^2, \quad \rho := (\sigma_{xy}/\sigma_x\sigma_y)^2 \leq 1, \quad -1 \leq \rho \leq +1. \qquad \textbf{[6]}$$

(iii) The extremal cases $r = \pm 1$ or $r^2 = 1$, have discriminant 0, that is $Q(\lambda)$ has a repeated real root, λ_0 say. But then $Q(\lambda_0)$ is the sum of squares of $\lambda_0(x_i - \bar{x}) + (y_i - \bar{y})$, which is zero. So each term is 0:

$$\lambda_0(x_i-\bar{x}) + (y_i-\bar{y}) = 0 \quad (i = 1\ldots n).$$

So all the points (x_i, y_i) lie on a line through $(\bar{x}, \bar{y})$ with slope $-\lambda_0$. **[4]**

Similarly, $\rho = \pm 1$ iff $Q(\lambda)$ has a repeated real root λ_0. Then

$$Q(\lambda_0) = E[(\lambda_0(x-Ex) + (y-Ey))^2] = 0.$$

So the random variable $\lambda_0(x - Ex) + (y - Ey)$ is zero (a.s. – except possibly on some set of probability 0). So all values of (x, y) lie on a straight line through the centroid (Ex, Ey) of slope $-\lambda_0$, a.s. **[3]**
(iv) From the strong law of large numbers, $\overline{xy} \to E[xy]$ a.s., $\overline{x^2} \to E[x^2]$ a.s., $\overline{y^2} \to E[y^2]$ a.s. So by (i), $r \to \rho$ a.s. **[2]**
[Seen – Problems]

Q3: *Bernoulli-Laplace urn.* (i) $(1+x)^{2d} \equiv (1+x)^d.(1+x)^d$. Equate coefficients of x^d left and right:

$$\binom{2d}{d} = \sum_{i=0}^{d} \binom{d}{i} . \binom{d}{d-i} = \sum_i \binom{d}{i}^2,$$

so the π_i in $HG(d)$ sum to 1, so $HG(d)$ is a probability distribution. **[4]**
(ii) A chain $P = (p_{ij})$ has *detailed balance (DB)* w.r.t. π if

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j. \qquad (DB)$$

Detailed balance is equivalent to reversibility (in time), and then the π in (DB) is the limit distribution of the chain. **[3]**
(iii) For $i \mapsto i+1$, one more black ball must go into I. So a white ($d-i$ of these) goes from I to II, and a black (again, $d-i$ of these) from II to I. Similarly, for $i+1 \mapsto i$, a black ($i+1$ of these) goes from I to II, and a white ($i+1$ of these) from II to I. For $i \mapsto i$, a black from I is interchanged with a black from II, or a white with a white (pr. $i(d-i)/d^2$ each). So

$$p_{i,i+1} = \Big(\frac{d-i}{d}\Big)^2, \quad p_{i,i-1} = \Big(\frac{i}{d}\Big)^2, \quad p_{i,i} = 2i(d-i)/d^2,$$
$$p_{ij} = 0 \ (j \neq i, i \pm 1). \qquad \textbf{[6]}$$

(iv) With $P = (p_{ij})$ as above and $\pi = HG(d)$,

$$\pi_i p_{i,i+1} = \frac{1}{\binom{2d}{d}} \binom{d}{i}^2 . \Big(\frac{d-i}{d}\Big)^2 = \frac{1}{\binom{2d}{d}} \binom{d-1}{i}^2,$$

and similarly

$$\pi_{i+1} p_{i+1,i} = \frac{1}{\binom{2d}{d}} \binom{d}{i+1}^2 . \Big(\frac{i+1}{d}\Big)^2 = \frac{1}{\binom{2d}{d}} \binom{d-1}{i}^2 = \pi_i p_{i,i+1},$$

proving DB, and so reversibility, with $\pi = HG(d)$ as limit distribution. **[6]**
(v) $\pi_i = 1/\mu_i$ by the Erdös-Feller-Pollard theorem, so $\mu_0 = 1/\pi_0 = \binom{2d}{d}$. **[1]**
(vi) By Stirling's formula,

$$\mu_0 \sim \frac{\sqrt{2\pi} e^{-2d} (2d)^{2d+\frac{1}{2}}}{(\sqrt{2\pi} e^{-d} d^{d+\frac{1}{2}})^2} = \frac{4^d}{\sqrt{\pi d}}.$$

Now as d is already very large (of the order of Avogadro's number 6×10^{23}), 4^d is astronomically vast – effectively infinite. This reconciles *microscopic reversibility* with *macroscopic irreversibility* in Statistical Mechanics. **[5]**
[Seen – Problems]

Q4: *Brownian motion.* (i) *Standard Brownian motion* $B = (B_t)$ on $\mathbb{R}$ ($BM(\mathbb{R})$, or BM) is defined to:
(a) start at 0, $B_0 := 0$; **[1]**
(b) have independent stationary Gaussian increments: $B_{s+t} - B_s \sim N(0, t)$ and independent of $\sigma(B_u : 0 \le u \le s)$; **[1,1,1]**
(c) have continuous paths: $t \mapsto B_t$ is continuous in t, a.s. **[1]**
(ii) *Mean and covariance.* The mean is 0 (by (b) with $s = 0$). **[1]**
So the covariance is, with $s \le t$,

$$cov(B_s, B_t) = E[B_s B_t] = E[B_s(B_s+(B_t-B_s)] = E[B_s^2]+E[B_s].E[B_t-B_s] = s+0.0 = s,$$

as $var(B_s) = E[B_s^2] = s$ and using independent increments. Similarly for $t \le s$. Combining,

$$E[B_t] = 0, \qquad cov(B_s, B_t) = \min(s, t). \qquad [\mathbf{5}]$$

(iii) *Scaling property.* For any $c > 0$, with B BM write

$$B_c(t) := c^{-1}B(c^2 t), \qquad t \ge 0.$$

Then B_c is Gaussian, with mean 0, variance $c^{-2} \times c^2 t = t$ and covariance

$$cov(B_c(s), B_c(t)) = c^{-2}E(B_c(s).B_c(t)) = c^{-2}\min(c^2 s, c^2 t)$$
$$= \min(s, t) = cov(B(s), B(t)).$$

Also B_c has continuous paths, as B does. So B_c has all the properties of Brownian motion. So, B_c *is* Brownian motion. **[4]**
(iv) *Local behaviour.* B_c is derived from B by *Brownian scaling* with *scale-factor* $c > 0$. As for each $u > 0$ $(B(ut) : t \ge 0) = (\sqrt{u}B(t) : t \ge 0)$ in law, B is called *self-similar* with *index* 1/2. Brownian motion is thus a *fractal.* A piece of Brownian path, looked at under a microscope, still looks Brownian, however much we 'zoom in and magnify' – unlike the functions f of calculus, which begin to look straight, as they have tangents. **[3]**
(v) *Brownian bridge* $X_t := B_t - tB_1$ ($t \in [0, 1]$) is Gaussian: it is obtained from the Gaussian process B by linear operations, and these preserve Gaussianity, by definition of the multivariate normal distribution. It has mean 0 (as B does), and covariance

$$E[X_s X_t] = E[(B_s - sB_1)(B_t - tB_1)] = E[B_s B_t] - tE[B_s B_1] - sE[B_t B_1] + stE[B_1^2]$$
$$= \min(s, t) - t\min(s, 1) - s\min(t, 1) + st = \min(s, t) - st - st + st = \min(s, t) - st.$$

[7]

[Seen – lectures and problems – apart from (iii)] NHB