PROBABILITY FOR STATISTICS: EXAMINATION SOLUTIONS 2015-16

Q1 *Chi-square distribution* with n *degrees of freedom*, $\chi^2(n)$.

This is defined as the distribution of $X_1^2+\ldots+X_n^2$, with X_i iid $N(0,1)$. **[3]**

(i) For $n=1$, the mean is 1, because a $\chi^2(1)$ is the square of a standard normal, and a standard normal has mean 0 and variance 1. The variance is 2, because the fourth moment of a standard normal X is 3, and $var(X^2) = E[(X^2)^2] - [E(X^2)]^2 = 3-1=2$. For general n, the mean is n because means add, and the variance is $2n$ because variances add over independent summands. **[3,3]**

(ii) For X standard normal, the MGF of its square X^2 is

$$M(t) := \int e^{tx^2}\phi(x)dx = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{tx^2}.e^{-\frac{1}{2}x^2}dx = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{1}{2}(1-2t)x^2}dx.$$

The integral converges only for $t<\frac{1}{2}$, when (normal integral) it is $1/\sqrt{(1-2t)}$:

$$M(t) = 1/\sqrt{1-2t} \qquad (t<\frac{1}{2}) \qquad \text{for } X\sim N(0,1). \qquad [\mathbf{4}]$$

So by definition of $\chi^2(n)$, the MGF of a $\chi^2(n)$ is

$$M(t) = 1/(1-2t)^{\frac{1}{2}n} \qquad (t<\frac{1}{2}) \qquad \text{for } X\sim \chi^2(n).$$

Replacing t by it by analytic continuation, the characteristic function is

$$\phi(t) = 1/(1-2it)^{\frac{1}{2}n}. \qquad [\mathbf{4}]$$

(iii) First, the required density $f(.)$ is a density, as $f\geq 0$ and $\int f = 1$:

$$\int f(x)dx = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}.\int_0^\infty x^{\frac{1}{2}n-1}\exp(-\frac{1}{2}x)dx = \frac{1}{\Gamma(\frac{1}{2}n)}.\int_0^\infty u^{\frac{1}{2}n-1}\exp(-u)du = 1$$

($u:=\frac{1}{2}x$), by definition of the Gamma function. **[4]**

Its MGF is

$$M(t) = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}.\int_0^\infty e^{tx}.x^{\frac{1}{2}n-1}\exp(-\frac{1}{2}x)dx = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}.\int_0^\infty x^{\frac{1}{2}n-1}\exp(-\frac{1}{2}x(1-2t))dx.$$

Substitute $u:=\frac{1}{2}x(1-2t)$ in the integral. This gives

$$M(t) = (1-2t)^{-\frac{1}{2}n}.\frac{1}{\Gamma(\frac{1}{2}n)}.\int_0^\infty u^{\frac{1}{2}n-1}e^{-u}du = (1-2t)^{-\frac{1}{2}n}, \quad \phi(t) = (1-2it)^{-\frac{1}{2}n}.$$

So it has the required MGF and CF, so is the required density. // **[4]**

[Seen – lectures, + Mock Exam 2012, Q1]

Q2 *Affine-invariance; skewness and kurtosis; testing for normality.*
(i) The cumulants κ_n are the coefficients in the cumulant generating function (CGF) $K(t)$ – log of the moment GF: $M(t) := \sum_0^\infty \mu_k t^n/n!$,
$K(t) := \log M(t) = \sum_0^\infty \kappa_n t^n/n!$. **[2]**
(ii) The *skewness* and *kurtosis* (parameters) are defined by ($\sigma^2 = \mu_{2,0}$)

$$\gamma_1 := \kappa_3/\kappa_2^{3/2} = \mu_{3,0}/\mu_{2,0}^{3/2} = \mu_{3,0}/\sigma^3, \qquad \gamma_2 := \kappa_4/\kappa_2^2 = \frac{\mu_{4,0}}{\sigma^4} - 3 = \frac{\mu_{4,0}}{\mu_{2,0}^2} - 3.$$

Their sample counterparts (statistics), the *sample skewness* $\hat{\gamma}_1$ and *sample kurtosis* $\hat{\gamma}_2$, are

$$\hat{\gamma}_1 := \hat{\mu}_{3,0}/\hat{\mu}_{2,0}^{3/2}, \qquad \hat{\gamma}_2 := \frac{\hat{\mu}_{4,0}}{\hat{\mu}_{2,0}^2} - 3. \qquad [\mathbf{3}], [\mathbf{3}]$$

(iii) By SLLN applied to X^k,

$$\hat{\mu}_k := \overline{X^k} \to E[X^k] = \mu_k \quad a.s. \quad (n \to \infty).$$

The kth *sample central moment* is $\hat{\mu}_k^0 := \overline{(X - \overline{X})^k}$. Then

$$\hat{\mu}_k^0 := \overline{(X - \overline{X})^k} = \overline{\sum\nolimits_0^k \binom{k}{i} X^i (-)^{k-i} (\overline{X})^{k-i}} = \sum\nolimits_0^k \binom{k}{i} (\overline{X^i})(-)^{k-i}(\overline{X})^{k-i}.$$

By SLLN, as $n \to \infty$ this tends a.s. to its population counterpart, as

$$\sum\nolimits_0^k \binom{k}{i} E[X^i](-)^{k-i}[EX]^{k-i} = E[(X - EX)^k] = \mu_k^0. \qquad // \ [\mathbf{6}]$$

(iv) So (from their definitions in terms of sample central moments) the sample skewness and sample kurtosis tend to their population counterparts also. **[2]**
(v) As the MGF of $N(\mu, \sigma)$ is $M(t) = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}$, its CGF is $K(t) = \mu t + \frac{1}{2}\sigma^2 t^2$. So the population is normal iff all cumulants higher than the second vanish. In particular, normal skewness and kurtosis vanish. **[2]**
(vi) Skewness and kurtosis, sample and population versions, are affine-equivariant: under $x \mapsto ax + b$, b drops out because of the *centring*, and a because of the *scaling* (a^3 in each of the numerator and denominator in the above). **[2]**
(vii) Sample skewness and kurtosis are suitable test statistics as they are affine-invariant, and so is the hypothesis of normality (with mean and variance unspecified). A test statistic of the form $a\hat{\gamma}_1^2 + b\hat{\gamma}_2^2$ will do, rejecting (in view of (iv) and (v)) if this is too big (details – below – not required). **[5]**
[Fisher: for normality, $\hat{\gamma}_1 \sim N(0, 6/n)$; $\hat{\gamma}_2^2 \sim N(0, 24/n)$; $n(\hat{\gamma}_1^2/6 + \hat{\gamma}_2^2/24) \sim \chi^2(2)$; Keeping, 8.18; cf. Jarque-Bera test.]
[(i) - (iii) and (vi): Seen, Problems 4; (iv), (v), (vii): unseen].

Q3 *Edgeworth's theorem.*
Definition. An n-vector X has an *n-variate normal* (or *Gaussian*) distribution iff $a^T X$ is univariate normal for all constant n-vectors a. **[2]**

If X is multivariate normal with mean vector μ and covariance matrix Σ, write $X \sim N(\mu, \Sigma)$. Then Σ is symmetric and non-negative definite ($\Sigma \geq 0$). If further Σ is positive definite ($\Sigma > 0$), we quote from Linear Algebra (spectral decomposition theorem) that Σ^{-1}, $\Sigma^{\frac{1}{2}}$ and $\Sigma^{-\frac{1}{2}}$ exist. **[2]**

Theorem (Edgeworth, 1893). If μ is an n-vector, $\Sigma > 0$ a symmetric positive definite $n \times n$ matrix, then
(i)

$$f(x) := \frac{1}{(2\pi)^{\frac{1}{2}n} |\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\}$$

is an n-dimensional prob. density function (of a random n-vector X, say); **[2]**
(ii) X is multinormal $N(\mu, \Sigma)$. **[2]**

Proof. Write $Y := \Sigma^{-\frac{1}{2}} X$ ($\Sigma^{-\frac{1}{2}}$ exists as $\Sigma > 0$, given). Then Y has covariance matrix $\Sigma^{-\frac{1}{2}} \Sigma (\Sigma^{-\frac{1}{2}})^T$. Since $\Sigma = \Sigma^T$ and $\Sigma = \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}$, Y has covariance matrix I (the components Y_i of Y are uncorrelated). **[2]**

Change variables as above, with $y = \Sigma^{-\frac{1}{2}} x$, so $x = \Sigma^{\frac{1}{2}} y$, and $\nu := \Sigma^{-\frac{1}{2}} \mu$, so $\mu = \Sigma^{\frac{1}{2}} \nu$. So

$$x - \mu = \Sigma^{\frac{1}{2}}(y - \nu), \quad (x-\mu)^T \Sigma^{-1} (x-\mu) = (y-\nu)^T \Sigma^{\frac{1}{2}} \Sigma^{-1} \Sigma^{\frac{1}{2}} (y-\nu) = (y-\nu)^T (y-\nu).$$

[2]

The Jacobian is (taking $A = \Sigma^{-\frac{1}{2}}$) $J = \partial x / \partial y = det(\Sigma^{\frac{1}{2}}), = (det \Sigma)^{\frac{1}{2}}$ by the product theorem for determinants. **[2]**

By the change of density formula, Y has density

$$g(y) = \frac{1}{(2\pi)^{\frac{1}{2}n} |\Sigma|^{\frac{1}{2}}} . |\Sigma|^{\frac{1}{2}} . \exp\{-\frac{1}{2}(y-\nu)^T (y-\nu)\}, \quad = \Pi_{i=1}^n \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\{-\frac{1}{2}(y_i - \nu_i)^2\}.$$

So the components Y_i are independent $N(\nu_i, 1)$. So Y is $N(\nu, I)$. **[3]**
(i) Taking $A = B = \mathbb{R}^n$, $\int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}^n} g(y) dy, = 1$ as g is a probability density, as above. So f is also a probability density. **[4]**
(ii) $X = \Sigma^{\frac{1}{2}} Y$ is a linear transf. of Y, so is multivariate normal as Y is.
$E[X] = \Sigma^{\frac{1}{2}} E[Y] = \Sigma^{\frac{1}{2}} \nu = \Sigma^{\frac{1}{2}} . \Sigma^{-\frac{1}{2}} \mu = \mu$, $cov(X) = \Sigma^{\frac{1}{2}} cov(Y) (\Sigma^{\frac{1}{2}})^T = \Sigma^{\frac{1}{2}} I \Sigma^{\frac{1}{2}} = \Sigma$. So X is multinormal $N(\mu, \Sigma)$. **[4]**
[Seen, lectures, L16]

Q4 *Tower property of conditional expectations; Conditional Mean Formula.*
(i) (Tower property).
If $\mathcal{C} \subset \mathcal{B}$, $E[E(Y|\mathcal{B})\ |\mathcal{C}] = E[Y|\mathcal{C}]$ a.s.
Proof. $E_{\mathcal{C}} E_{\mathcal{B}} Y$ is $\mathcal{C}$-measurable, and for $C \in \mathcal{C} \subset \mathcal{B}$,

$$\int_C E_{\mathcal{C}}[E_{\mathcal{B}} Y] dP = \int_C E_{\mathcal{B}} Y dP \qquad \text{(definition of } E_{\mathcal{C}} \text{ as } C \in \mathcal{C})$$

$$= \int_C Y dP \qquad \text{(definition of } E_{\mathcal{B}} \text{ as } C \in \mathcal{B}).$$

So $E_{\mathcal{C}}[E_{\mathcal{B}} Y]$ satisfies the defining relation for $E_{\mathcal{C}} Y$. Being also $\mathcal{C}$-measurable, it *is* $E_{\mathcal{C}} Y$ (a.s.). // **[10]**
(i') (Tower property 'the other way round').
If $\mathcal{C} \subset \mathcal{B}$, $E[E(Y|\mathcal{C})\ |\mathcal{B}] = E[Y|\mathcal{C}]$ a.s. *Proof.* $E[Y|\mathcal{C}]$ is $\mathcal{C}$-measurable, so $\mathcal{B}$-measurable as $\mathcal{C} \subset \mathcal{B}$, so $E[.|\mathcal{B}]$ has no effect. // **[3]**
(ii) (Conditional expectation as projection).
By the tower property (either way round),

$$E[E[Y|\mathcal{C})]|\mathcal{C}] = E[Y|\mathcal{C}] \quad a.s.$$

So the operation $E[.|\mathcal{C}]$ is linear and *idempotent* (doing it twice is the same as doing it once), so is a *projection.* **[6]**
(iii) (Conditional Mean Formula).
Take $\mathcal{C}$ the trivial σ-field $\{\emptyset, \Omega\}$. This contains no information, so an expectation conditioning on it is the same as an unconditional expectation. The first form of the tower property now gives

$$E[E[X|\mathcal{B}]] = E[E[X|\mathcal{B}]\ |\{\emptyset, \Omega\}] = E[X|\{\emptyset, \Omega\}] = E[X] :$$

$$E[E[X|\mathcal{B}]] = E[X]. \qquad [\mathbf{6}]$$

[Seen, Problems, Prob 9 Q1, Prob 10 Q1]

NHB