McGILL UNIVERSITY

FACULTY OF SCIENCE

FINAL EXAMINATION

MATHEMATICS AND STATISTICS MATH 324

STATISTICS

Examiner: Professor A. C. Vandal
Associate examiner: Professor W. Anderson

Date: 18 April 2006
Time: 14:00–17:00

INSTRUCTIONS

- **Attempt all questions.**
- **Answer in the exam booklet(s) supplied.**
- **This exam will be marked out of 90, out of a possibility of 100 marks.**
- **The function** $\log x$ **is understood to mean** $\log_e(x) = \ln(x)$.
- **The expression** $\mathbb{1}[\textit{statement}]$ **takes on value 1 if** *statement* **is true and 0 if** *statement* **is false.**
- **Translation dictionaries are allowed.** Regular dictionaries are not permitted.
- **You are not allowed any notes, textbooks or similar material.**
- **Non-programmable calculators are allowed and expected.**
- **Good luck!**

This exam comprises the cover, 8 questions on 4 pages; 3 pages listing pmf's, pdf's and some of their properties; and 2 pages with tables of common sampling distributions.

1. [12 marks total]
Let $X_1, \ldots, X_n$ be a random sample with cdf F. For some $x \in \mathbb{R}$, let

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[X_i \leq x].$$

(a) [2 marks]
Show that $\hat{F}(x)$ is $F(x)$-unbiased.

(b) [4 marks]
Show that $\hat{F}(x)$ is $F(x)$-consistent. Be clear about any result you invoke to do so.

(c) [6 marks]
Show that

$$\frac{\sqrt{n}(\hat{F}(x) - F(x))}{\sqrt{F(x)(1-F(x))}} \xrightarrow{\mathcal{D}} Z \sim N(0,1),$$

where "$\xrightarrow{\mathcal{D}}$" means "converges in distribution to". Be clear about any result you invoke to do so.

2. [15 marks total]
Let $X_1, \ldots, X_n$ be a random sample from an $\text{Exp}(\theta)$ distribution with $\mathbb{E}[X_i] = \theta$, $i = 1, \ldots, n$.

(a) [4 marks]
Show that the maximum likelihood estimator of θ is $\hat{\theta} = \overline{X}$.

(b) [3 marks]
Show that $\overline{X} \sim \Gamma(n, \theta/n)$.

(c) [4 marks]
Assume that $n = 8$. Find $\text{P}\left[\hat{\theta} > 2\theta\right]$. (*Hint*: Find the distribution of $\frac{2n\overline{X}}{\theta}$.)

(d) [4 marks]
Let $\hat{F}_\theta(x)$ be the maximum likelihood estimator of the $\text{Exp}(\theta)$ cumulative distribution function (cdf) $F_\theta(x)$. Find an expression for $\text{Var}\left[\hat{F}_\theta(x)\right]$. (*Hint*: Use the fact that for any random variable Y with a moment-generating function $M_Y(t)$, $\mathbb{E}[\exp(tY)] = M_Y(t)$ whenever the expectation exists.)

CONTINUED

3. [12 marks total]
Let $Y_1, \ldots, Y_n$ be a random sample from a Poisson(λ) distribution, $\lambda > 0$. Use the Cramér-Rao inequality to show that $\hat{\lambda} = \overline{Y}$ is the best (minimum variance) unbiased estimator of λ.

4. [12 marks total]
The following data consist of height measurements in centimeters of 5 randomly selected male children at ages 3 and 5.

Child	Height at 3 years (x_i)	Height at 5 years (y_i)	Difference ($y_i - x_i$)
1	95.6	114.8	19.2
2	75.4	84.0	8.6
3	76.6	99.7	23.1
4	99.4	121.0	21.6
5	65.7	80.7	15.0
Observed sample mean	82.5	100.0	17.5
Observed sample variance	206.2	322.2	34.1

Assume that the data are observed from Normal populations.

(a) [5 marks]
Find a 95% confidence interval of the form $(0, b)$ for the variance of the height of male children at 3 years.

(b) [7 marks]
Find a 95% confidence interval of the form (a, b) with $-\infty < a < b < +\infty$ for the difference in height amongst boys between ages 3 and 5.

5. [10 marks total]
In a Bayesian setting, assume that the underlying random sample of height differences $y_i - x_i$ from **4.** given Δ have a $N(\Delta, 50)$ distribution. Assume that $\Delta \sim N(20, 34.1)$. Find a 90% credible interval for Δ.

CONTINUED

6. [12 marks total]
Let $X_1, \ldots, X_n$ be a random sample from a distribution with probability density function $f_\theta(x) = \theta x^{\theta-1}$ for $0 \leq x \leq 1$ and with $\theta > 0$ unknown.

(a) [4 marks]
Let $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, where θ_0 and θ_1 are given and $0 < \theta_0 < \theta_1$. Show that the rejection region of the most powerful test of level α has the form

$$\frac{1}{n}\sum_{i=1}^{n} \log X_i > k$$

for some k (that depends on α). You do not need identify k at this stage.

(b) [4 marks]
Obtain an approximate value for k in part **(a)** expressed in terms of n, θ_0 and z_α, where z_α satisfies $\mathrm{P}\left[Z > z_\alpha\right] = \alpha$ when $Z \sim N(0,1)$.

(c) [4 marks]
With the same random sample as above, we now wish to test $H_0 : \theta = \theta_0$ against $H_1' : \theta \neq \theta_0$. Obtain a test statistic with a χ^2_ν distribution under H_0 and an approximate rejection region for this test statistic at a level of $\alpha = 0.05$.

7. [12 marks total]
Consider the data from Question **4**. Let Y_i be the height of child i at 5 years, and x_i be the height of child i at age 3.

(a) [5 marks]
Use the observations to obtain least-squares estimates of α and β in the model

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

for $\epsilon_i \sim N(0, \sigma^2)$, $\sigma^2 > 0$ unknown.

(b) [7 marks]
Test $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ at the 5% level.

CONTINUED

8. [15 marks total; 1.5 mark per question]
Answer the following by true (T) or false (F). Do not justify your answer.

(a) True or False: If L and U are random variables such that a $1-\alpha$ confidence interval for real parameter θ is given by $[L, U]$, then $\mathrm{P}[L \leq \theta \leq U] = 1-\alpha$.

(b) True or False: If $X_1, \ldots, X_n$ is a $N(\mu, \sigma^2)$ random sample with μ and σ^2 known, then $\sqrt{n}\dfrac{\overline{X}-\mu}{\sqrt{S^2}} \sim t_n$.

(c) In a Bayesian setting, let $X_1, \ldots, X_n | \mu \sim N(\mu, \sigma^2)$ and $\mu \sim N(\nu, \tau^2)$, for $\sigma^2, \tau^2 > 0$ and $\nu \in \mathbb{R}$ fixed.
True or False: The posterior expectation of μ given $X_1 = x_1, \ldots, X_n = x_n$ converges to $\overline{x}$ as $n \to \infty$.

(d) Let $Y_1, \ldots, Y_n$ be a random sample from a $\Gamma(\alpha, \beta)$ distribution, and let $S^2 = \frac{1}{n}\sum_{i=1}(Y_i - \overline{Y})^2$ as usual.
True or False: $\mathbb{E}[S^2] = \alpha\beta^2$.

(e) True or False: A biased estimator can have a lower mean-squared error than a minimum variance unbiased estimator.

(f) Let $X_1, \ldots, X_n$ be a random sample from an $\text{Exp}(\theta)$ distribution with expectation θ. Let $\hat{\theta}_1 = nX_{\min}$ and $\hat{\theta}_2 = \overline{X}$.
True or False: The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is $\text{rel}(\hat{\theta}_1, \hat{\theta}_2) = n$.

(g) Let $Y_1, \ldots, Y_n$ be a random sample from $N(0, \sigma^2)$, $\sigma^2 > 0$ unknown. Define $\chi^2_{\nu,\gamma}$ as the real value that satisfies $\mathrm{P}\left[W > \chi^2_{\nu,\gamma}\right] = \gamma$, where $W \sim \chi^2_\nu$.
True or False: The most powerful α-level test of $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$ has the rejection $\sum_{i=1}^n X_i^2 > k_1$ for some k_1 if $\sigma_1^2 > \sigma_0^2$, or $\sum_{i=1}^n X_i^2 < k_2$ for some k_2 if $\sigma_1^2 < \sigma_0^2$.

(h) Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ unknown and $\sigma^2 > 0$ known.
True or False: The sample mean and sample variance of the random sample are independent.

(i) Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ both unknown.
True or False: The sample mean and sample variance of the random sample are independent.

(j) True or False: In a simple linear regression setting with normality, independence and homoscedasticity of errors, the slope estimator has a normal distribution.

END OF FINAL EXAMINATION.
A list of pmf's and pdf's follows.

List of probability mass functions and probability density functions:

- Beta **Beta**

  - If $X \sim \text{Beta}(\alpha, \beta)$, then X has probability density function

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}[0 < x < 1]$$

  for $\alpha, \beta > 0$. See Gamma distribution for properties of the Gamma function.

  - $\mathbb{E}[X] = \dfrac{\alpha}{\alpha+\beta}$.
  - $\text{Var}[X] = \dfrac{\alpha\beta}{(\alpha+\beta+1)^2(\alpha+\beta+2)}$.

- Bin **Binomial**

  - If $X \sim \text{Bin}(n, p)$, then X has probability function

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}\mathbb{1}[x \in \{0, 1, \ldots, n\}],$$

  where $n = 1, 2, \ldots$ and $0 < p < 1$.

  - $\mathbb{E}[X] = np$.
  - $\text{Var}[X] = np(1-p)$.
  - MGF: $M_X(t) = [1 - p + p\exp(t)]^n$.

- χ^2_ν **Chi-squared**

  - If $X \sim \chi^2_\nu$, then X has probability density function given by

$$f(x) = \frac{x^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)}\exp(-x/2)\mathbb{1}[x > 0].$$

  with $\nu > 0$.

  - See Gamma distribution for properties, as $X \sim \chi^2_\nu \Leftrightarrow X \sim \Gamma\left(\frac{\nu}{2}, 2\right)$.

- Exp **Exponential**

  - If $X \sim \text{Exp}(\beta)$, then X has probability density function

$$f(x) = \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right)\mathbb{1}[x > 0]$$

  for $\beta > 0$.

- CDF: $F(x) = \begin{cases} 1 - \exp(-x/\beta) & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$
- See Gamma distribution for other properties, as $X \sim \text{Exp}(\beta) \Leftrightarrow X \sim \Gamma(1, \beta)$.

-Γ **Gamma**

- If $X \sim \Gamma(\alpha, \beta)$, then X has probability density function

$$f(x) = \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} \exp(-x/\beta) \mathbb{1}[x > 0]$$

for $\alpha, \beta > 0$, where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx$. The Gamma function $\Gamma(\alpha)$ has the following properties:
  - $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$;
  - $\Gamma(1) = 1$;
  - If n is a positive integer, $\Gamma(n + 1) = n!$;
  - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
- $\mathbb{E}[X] = \alpha\beta$.
- $\text{Var}[X] = \alpha\beta^2$.
- MGF: $M_X(t) = (1 - \beta t)^{-\alpha}$.

- Geom **Geometric**

- If $X \sim \text{Geom}(p)$, then X has probability mass function

$$f(x) = p(1 - p)^x$$

for $p \in [0, 1]$ and $x \in 0, 1, 2, \ldots$.
- $\mathbb{E}[X] = \dfrac{1-p}{p}$.
- $\text{Var}[X] = \dfrac{1-p}{p^2}$.
- MGF: $M_X(t) = \frac{p}{1-(1-p)e^t}$.

- N **Normal**

- If $X \sim \text{N}(\mu, \sigma^2)$, then X has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right] \mathbb{1}[x \in \mathbb{R}]$$

for $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

- $\mathbb{E}[X] = \mu$.
- $\text{Var}[X] = \sigma^2$.
- MGF: $M_X(t) = \exp\left(\mu t + \sigma^2 \frac{t^2}{2}\right)$.

- NB **Negative Binomial**

- If $X \sim \text{NB}(r, p)$, then X has probability function

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{if } x = r, r+1, r+2, \dots \\ 0 & \text{otherwise,} \end{cases}$$

for $r = 1, 2, \dots,$ and $0 < p < 1$.

- $\mathbb{E}[X] = \dfrac{r}{p}$.
- $\text{Var}[X] = \dfrac{r(1-p)}{p^2}$.
- MGF: $M_X(t) = \left(\dfrac{pe}{1-(1-p)e^t}\right)^r$.

- Po **Poisson**

- If $X \sim \text{Po}(\lambda)$, then X has probability function

$$f(x) = \exp(-\lambda)\frac{\lambda^x}{x!}\mathbb{1}[x \in \{0, 1, 2, \dots\}]$$

for $\lambda > 0$.

- $\mathbb{E}[X] = \lambda$.
- $\text{Var}[X] = \lambda$.
- MGF: $M_X(t) = \exp[\lambda(e^t - 1)]$.

END OF LIST OF PMF's and PDF's
Tables of some sampling distributions follow.

Upper tail probabilities of the standard Normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

t distribution quantiles

Degrees of freedom	Upper tail probability 0.1	0.05	0.025	0.01	0.005
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.05
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.79
26	1.31	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
35	1.31	1.69	2.03	2.44	2.72
40	1.30	1.68	2.02	2.42	2.70
45	1.30	1.68	2.01	2.41	2.69
50	1.30	1.68	2.01	2.40	2.68
55	1.30	1.67	2.00	2.40	2.67
60	1.30	1.67	2.00	2.39	2.66
80	1.29	1.66	1.99	2.37	2.64
100	1.29	1.66	1.98	2.36	2.63
120	1.29	1.66	1.98	2.36	2.62
∞	1.28	1.64	1.96	2.33	2.58

χ^2 distribution quantiles

Degrees of freedom	Upper tail probability 0.1	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.95
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67
35	46.06	49.80	53.20	57.34	60.27
40	51.81	55.76	59.34	63.69	66.77
45	57.51	61.66	65.41	69.96	73.17
50	63.17	67.50	71.42	76.15	79.49
55	68.80	73.31	77.38	82.29	85.75
60	74.40	79.08	83.30	88.38	91.95
80	96.58	101.88	106.63	112.33	116.32
100	118.50	124.34	129.56	135.81	140.17
120	140.23	146.57	152.21	158.95	163.65
∞					