Tentamen
MVE302 Sannolikhet och statistik

2019-06-05 kl. 8:30 - 12:30

Examinator: Johan Jonasson, Matematiska vetenskaper, Chalmers

Telefonvakt: Oskar Allerbo, telefon: 031-7725325

Hjälpmedel: Valfri miniräknare. Två blad (dvs fyra sidor) handskrivna anteckningar. Tabeller finns längst bak på tentamenstesen.

Denna tentamen utgör grunden för betygssättning. För betyg 3 krävs minst 16 poäng, för betyg 4 minst 24 poäng och för betyg 5 minst 32 poäng

---

1. (6p) Paret (X, Y) av stokastiska variabler har täthetsfunktion $f(x, y) = c(x - xy + y)$, $0 < x < 1$, $0 < y < 1$. Bestäm konstanten c, tätheterna f_X och f_Y för X och Y, väntevärdena $\mathbb{E}[X]$ och $\mathbb{E}[Y]$ och $\mathbb{C}\mathrm{ov}(X, Y)$.

2. (5p) En rättvis slant har singlats tre gånger och man har erhållit ett antal klave som vi betecknar med X. Sedan, givet $X = x$, har det dragits en Poissonfördelad stokastisk variabel Y med väntevärde x. Om du observerar $Y = 2$, vad är den betingade sannolikheten att $X = x$, $x = 0, 1, 2, 3$?

3. (6p) Gör en linjär regression med följande data:

x	1	2	3	4	5
y	96	84	70	58	52

Skatta regressionslinjen $y = a + bx$, gör ett symmetriskt konfidensintervall med 95% konfidensgrad för b och gör ett symmetriskt prediktionsintervall med prediktionsgrad 90% för en ny observation vid $x = 6$.

4. Ett givet flygplan tar 100 passagerare.

(a) (3p) Man vet av erfarenhet att i genomsnitt 5% av dem som köpt biljetter uteblir från sin flight. Därför kan flygbolaget sälja en aning mer än 100 biljetter. Hur många biljetter kan flygbolaget sälja om man vill vara minst 95% säker på inte mer än 100 passagerare ska komma till flighten?

(b) (3p) Om vi för enkelhets skull antar att det kommer exakt 100 passagerare till flighten varav exakt hälften kvinnor och hälften män, vad är sannolikheten att passagerarnas samlade vikt överstiger 7.5 ton? En på måfå vald kvinna kan antas ha en vikt som är normalfördelad med väntevärde 63 och varians 10^2 och detsamma för en på måfå vald man om siffrorna byts till väntevärde 83 och varians 15^2.

5. (5p) Betrakta täthetsfunktionen

$$f(x) = \theta^2 x e^{-\theta x},\ x > 0.$$

Här är $\theta > 0$ en okänd parameter. Ett stickprov $X_1, X_2, \ldots, X_n$ på denna fördelning samlas in. Vad blir ML-skattningen av θ?

6. (5p) Jensens olikhet:

(a) (3p) Låt $f(x)$ vara en funktion som är två gånger deriverbar med $f''(x) > 0$ för alla x. Visa att det för en stokastisk variabel sådan att både $\mathbb{E}[X]$ och $E[f(X)]$ existerar, gäller att

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

(b) (2p) Använd resultatet i (a) till att visa att $\ln \mathbb{E}[e^X] \geq \mathbb{E}[X]$ (förutsatt att $\mathbb{E}[e^X]$ existerar). Glöm inte att du kan göra denna del även om du inte löser (a).

Tips för (a): Skriv $\mu = \mathbb{E}[X]$ och observera att tangenten till f i μ ligger under kurvan $f(x)$.

7. Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov på en normalfördelning med okänt väntevärde μ och okänd varians σ^2.

(a) (2p) Gör ett nedåt begränsat 95% konfidensintervall för σ^2.

(b) (2p) Hur stort måste n vara för att styrkan i testet $H_0 : \sigma^2 = 1$ mot $H_A : \sigma^2 > 1$ på 5% signifikansnivå ska vara minst 95% för $\sigma^2 = 4$?

8. (3p) Låt X vara en stokastisk variabel som kan anta värdena 0, 1 eller 2 och låt M_X vara dess momentgenererande funktion. Bestäm p_k, $k = 0, 1, 2$, då $M_X(1) = 1/6 + e/2 + e^2/3$ och $M_X(2) = 1/6 + e^2/2 + e^4/3$.

Lycka till!
Johan Jonasson

Tabell 1: Values of the cdf $\Phi(x)$ of the standard normal distribution [e.g., $\Phi(1.41) = 0.921$]

x	0	1	2	3	4	5	6	7	8	9
0.0	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1	.540	.544	.548	.552	.556	.560	.564	.568	.571	.575
0.2	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5	.692	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6	.726	.729	.732	.736	.739	.742	.745	.749	.752	.755
0.7	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813
0.9	.816	.819	.821	.824	.826	.829	.832	.834	.836	.839
1.0	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.892	.894	.896	.898	.900	.902
1.3	.903	.905	.907	.908	.910	.912	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932
1.5	.933	.934	.936	.937	.938	.939	.941	.942	.943	.944
1.6	.945	.946	.947	.948	.950	.951	.952	.952	.9545	.954
1.7	.955	.956	.957	.958	.959	.960	.961	.962	.962	.963
1.8	.964	.965	.966	.966	.967	.968	.969	.969	.970	.971
1.9	.971	.972	.973	.973	.974	.974	.975	.976	.976	.977
2.0	.977	.978	.978	.979	.979	.980	.980	.981	.981	.982
2.1	.982	.983	.983	.983	.984	.984	.985	.985	.985	.986
2.2	.986	.986	.987	.987	.988	.988	.988	.988	.989	.989
2.3	.989	.990	.990	.990	.990	.991	.991	.991	.991	.992
2.4	.992	.992	.992	.992	.993	.993	.993	.993	.993	.994
2.5	.994	.994	.994	.994	.995	.995	.995	.995	.995	.995
2.6	.995	.996	.996	.996	.996	.996	.996	.996	.996	.996
2.7	.996	.997	.997	.997	.997	.997	.997	.997	.997	.997
2.8	.997	.998	.998	.998	.998	.998	.998	.998	.998	.998
2.9	.998	.998	.998	.998	.998	.998	.998	.998	.999	.999

Tabell 2: Values of $\Phi(x)$ commonly used in confidence intervals and tests, and the corresponding x values

$\Phi(x)$	0.90	0.95	0.975	0.99	0.995
x	1.28	1.64	1.96	2.33	2.58

Tabell 3: Percentiles of the t distribution with DF degrees of freedom [e.g., $F_{t_7}(1.89) = 0.95$]

DF	0.95	0.975	0.99	0.995	DF	0.95	0.975	0.99	0.995
1	6.31	12.71	31.82	63.66	16	1.75	2.12	2.58	2.92
2	2.92	4.30	6.96	9.92	17	1.74	2.11	2.58	2.90
3	2.35	3.18	4,54	5.84	18	1.73	2.10	2.55	2.88
4	2.13	2.78	3.74	4.60	19	1.73	2.09	2.54	2.86
5	2.02	2.57	3.36	4.03	20	1.72	2.09	2.53	2.85
6	1.94	2.45	3.14	3.71	21	1.72	2.08	2.52	2.83
7	1.89	2.36	3.00	3.50	22	1.72	2.07	2.51	2.82
8	1.86	2.31	2.90	3.36	23	1.71	2.07	2.50	2.81
9	1.83	2.26	2.82	3.25	24	1.71	2.06	2.49	2.80
10	1.81	2.23	2.76	3.17	25	1.71	2.06	2.49	2.79
11	1.80	2.20	2.72	3.11	26	1.71	2.06	2.48	2.78
12	1.78	2.18	2.68	3.05	27	1.70	2.05	2.47	2.77
13	1.77	2.16	2.65	3.01	28	1.70	2.05	2.47	2.76
14	1.76	2.14	2.62	2.98	29	1.70	2.05	2.46	2.76
15	1.75	2.13	2.60	2.95	30	1.70	2.04	2.46	2.75

Tabell 4: Percentiles of the chi-square distribution with DF degrees of freedom [e.g., $F_{\chi^2_{20}}(10.85) = 0.05$]

DF	0.025	0.05	0.95	0.975	DF	0.025	0.05	0.95	0.975
1	0.001	0.004	3.84	5.02	16	6.91	7.96	26.30	28.84
2	0.05	0.10	5.99	7.38	17	7.56	8.67	27.59	30.19
3	0.22	0.35	7.82	9.34	18	8.23	9.39	28.87	31.53
4	0.48	0.71	9.49	11.14	19	8.91	10.12	30.14	32.85
5	0.83	1.14	11.07	12.83	20	9.59	10.85	31.41	34.17
6	1.24	1.64	12.59	14.45	21	10.28	11.60	32.67	35.48
7	1.69	2.17	14.07	16.01	22	10.98	12.34	33.92	36.78
8	2.18	2.73	15.51	17.54	23	11.69	13.09	35.17	38.08
9	2.70	3.32	19.92	19.02	24	12.40	13.85	36.42	39.36
10	3.25	3.94	18.31	20.48	25	13.12	14.61	37.65	40.65
11	3.82	4.58	19.68	21.92	26	13.84	15.38	38.88	41.92
12	4.40	5.23	21.03	23.34	27	14.57	16.15	40.11	43.19
13	5.01	5.89	22.36	27.74	28	15.31	16.93	41.34	44.46
14	5.63	6.57	23.68	26.12	29	16.05	17.71	42.56	45.72
15	6.26	7.26	25.00	27.49	30	16.79	18.49	43.77	46.98

Tabell 5: Percentiles of the F distribution with r and s degrees of freedom [e.g., $F_{F_{8,20}}(2.45) = 0.95$]

	2.5 % percentile								
s	$r = 2$	3	4	5	6	7	8	9	10
2	0.026	0.062	0.094	0.119	0.138	0.153	0.165	0.175	0.183
3	0.026	0.065	0.100	0.129	0.152	0.170	0.185	0.197	0.207
4	0.025	0.066	0.104	0.135	0.161	0.181	0.198	0.212	0.224
5	0.025	0.067	0.107	0.140	0.167	0.189	0.208	0.223	0.236
6	0.025	0.068	0.109	0.143	0.172	0.195	0.215	0.231	0.246
7	0.025	0.068	0.110	0.146	0.176	0.200	0.221	0.238	0.253
8	0.025	0.069	0.111	0.148	0.179	0.204	0.226	0.244	0.259
9	0.025	0.069	0.112	0.150	0.181	0.207	0.230	0.248	0.265
10	0.025	0.069	0.113	0.151	0.183	0.210	0.233	0.252	0.269
12	0.025	0.070	0.114	0.153	0.186	0.214	0.238	0.259	0.276
15	0.025	0.070	0.116	0.156	0.190	0.219	0.244	0.265	0.284
16	0.025	0.070	0.116	0.156	0.191	0.220	0.245	0.267	0.286
18	0.025	0.070	0.116	0.157	0.192	0.222	0.248	0.270	0.290
20	0.025	0.071	0.117	0.158	0.193	0.224	0.250	0.273	0.293
21	0.025	0.071	0.117	0.158	0.194	0.225	0.251	0.274	0.294
24	0.025	0.071	0.117	0.159	0.195	0.227	0.253	0.277	0.297
25	0.025	0.071	0.118	0.160	0.196	0.227	0.254	0.278	0.298
27	0.025	0.071	0.118	0.160	0.197	0.228	0.255	0.279	0.300
28	0.025	0.071	0.118	0.160	0.197	0.228	0.256	0.280	0.301
30	0.025	0.071	0.118	0.161	0.197	0.229	0.257	0.281	0.302

	95 % percentile								
s	$r = 2$	3	4	5	6	7	8	9	10
2	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
12	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
15	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
18	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
20	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
24	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
27	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
30	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16

s	$r=2$	3	4	5	6	7	8	9	10
	97.5 % percentile								
2	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
3	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
4	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
5	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
6	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
7	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
8	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30
9	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
10	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
12	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37
15	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
16	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99
18	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87
20	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77
21	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73
24	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64
25	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61
27	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57
28	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55
30	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51

Tabell 6: Critical values c for the Wilcoxon signed rank test, where n is the sample size and $C = n(n+1) - c$ [e.g., if $n = 20$, then $P(W \leq 61) = P(W \geq 149) \approx 0.05$]

n	0.025	0.05	$n(n+1)/2$	n	0.025	0.05	$n(n+1)/2$
5	0	1	15	18	41	48	171
6	1	3	21	19	47	54	190
7	3	4	28	20	53	61	210
8	4	6	36	21	59	68	231
9	6	9	45	22	67	76	253
10	9	11	55	23	74	84	276
11	11	14	66	24	82	92	300
12	14	18	78	25	90	101	325
13	18	22	91	26	99	111	351
14	22	26	105	27	108	120	378
15	26	31	120	28	117	131	406
16	30	36	136	29	127	141	435
17	35	42	153	30	138	152	465

Tabell 7: Critical values c for the Wilcoxon rank sum test, where m is the size of the smaller sample, and $C = m(m+n+1) - c$ [e.g., if $m = 4$ and $n = 8$, then $P(W \leq 16) = P(W \geq 36) \approx 0.05$]

n	$P(W \leq c)$	$m = 2$	3	4	5	6	7	8	9	10	11
2	0.025	3									
	0.05	3									
3	0.025	3	3								
	0.05	6	7								
4	0.025	3	6	11							
	0.05	3	7	12							
5	0.025	3	7	12	18						
	0.05	4	8	13	20						
6	0.025	3	8	13	19	27					
	0.05	4	9	14	21	29					
7	0.025	3	8	14	21	28	37				
	0.05	4	9	15	22	30	40				
8	0.025	4	9	15	22	30	39	50			
	0.05	5	10	16	24	32	42	52			
9	0.025	4	9	15	23	32	41	52	63		
	0.05	5	11	17	25	34	44	55	67		
10	0.025	4	10	16	24	33	43	54	66	79	
	0.05	5	11	18	27	36	46	57	70	83	
11	0.025	5	10	17	25	35	45	56	69	82	97
	0.05	5	12	19	28	38	48	60	73	87	101

Tentamen
MVE302 Sannolikhet och statistik

2019-06-05 kl. 8:30 - 12:30

Examinator: Johan Jonasson, Matematiska vetenskaper, Chalmers

Telefonvakt: Oskar Allerbo, telefon: 031-7725325

Hjälpmedel: Valfri miniräknare. Två blad (dvs fyra sidor) handskrivna anteckningar. Tabeller finns längst bak på tentamenstesen.

Denna tentamen utgör grunden för betygssättning. För betyg 3 krävs minst 16 poäng, för betyg 4 minst 24 poäng och för betyg 5 minst 32 poäng.

---

1. (6p) Paret (X, Y) av stokastiska variabler har täthetsfunktion $f(x, y) = c(x - xy + y)$, $0 < x < 1$, $0 < y < 1$. Bestäm konstanten c, tätheterna f_X och f_Y för X och Y, väntevärdena $\mathbb{E}[X]$ och $\mathbb{E}[Y]$ och $\mathbb{C}\mathrm{ov}(X, Y)$.

Lösning: Integralen av $c(x - xy + y)$ ska vara 1 och

$$\int_0^1 \int_0^1 (x - xy + y)dy\,dx = \int_0^1 \left(\frac{1}{2}x + \frac{1}{2}\right) dx = \frac{3}{4}$$

vilket betyder att $c = 4/3$. Vi har då

$$f_X(x) = \frac{4}{3} \int_0^1 (x - xy + y)dy$$

vilket vi ur ovanstående räkning läser ut är

$$f_X(x) = \frac{2}{3}(x + 1).$$

Eftersom den bivariata tätheten är helt symmetrisk i X och Y följer att $f_Y(y) = (2/3)(y + 1)$. Vidare är

$$\mathbb{E}[X] = \int_0^1 xf(x)dx = \frac{2}{3} \int_0^1 (x^2 + x)dx = \frac{2}{3}\left(\frac{1}{3} + \frac{1}{2}\right) = \frac{5}{9}$$

och tack vare symmetrin är $\mathbb{E}[Y] = 5/9$. För att få kovariansen behöver vi $\mathbb{E}[XY]$ och

$$\mathbb{E}[XY] = \frac{4}{3} \int_0^1 \int_0^1 xy(x - xy + y)dy\,dx$$

vilket efter en del räkning blir $8/27$. Därmed är

$$\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{8}{27} - \left(\frac{5}{9}\right)^2 = -\frac{1}{81}.$$

2. (5p) En rättvis slant har singlats tre gånger och man har erhållit ett antal klave som vi betecknar med X. Sedan, givet $X = x$, har det dragits en Poissonfördelad stokastisk variabel Y med väntevärde x. Om du observerar $Y = 2$, vad är den betingade sannolikheten $\mathbb{P}(X = x|Y = 2)$, $x = 0, 1, 2, 3$?

Lösning: För Poissonfördelningen är parametern lika med väntevärdet, så givet $X = x$ är $Y \sim Poi(x)$. Bayes formel ger

$$\begin{aligned}\mathbb{P}(X = x|Y = 2) &= \frac{\mathbb{P}(Y = 2|X = x)\mathbb{P}(X = x)}{\sum_{j=0}^{3} \mathbb{P}(Y = 2|X = j)\mathbb{P}(X = j)} \\ &= \frac{e^{-x}\frac{x^2}{2!}\binom{3}{x}\left(\frac{1}{2}\right)^3}{\sum_{j=0}^{3} e^{-j}\frac{j^2}{2!}\binom{3}{j}\left(\frac{1}{2}\right)^3} \\ &= \frac{\binom{3}{x}e^{-x}x^2}{\sum_{j=0}^{3}\binom{3}{j}e^{-j}j^2}\end{aligned}$$

3. (6p) Gör en linjär regression med följande data:

x	1	2	3	4	5
y	96	84	70	58	52

Skatta regressionslinjen $y = a + bx$, gör ett symmetriskt konfidensintervall med 95% konfidensgrad för b och gör ett symmetriskt prediktionsintervall med prediktionsgrad 90% för en ny observation vid $x = 6$.

Lösning: Data ger oss $n = 5$, $\sum x_k = 15$, $\sum y_k = 360$, $\sum x_k y_k = 966$, $\sum x_k^2 = 55$, $\sum y_k^2 = 27240$. Detta ger $S_{xx} = \sum x_k^2 - (1/5)(\sum x_k)^2 = 55 - (1/5)15^2 = 10$, $S_{xy} = \sum x_k y_k - (1/5)(\sum x_k)(\sum y_k) = 966 - (1/5) \cdot 15 \cdot 360 = -114$, $S_{yy} = 27240 - (1/5) \cdot 360^2 = 1320$. Detta ger i sin tur

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = -\frac{114}{10} = -11.4$$

och

$$\hat{a} = \overline{y} - \hat{b}\overline{x} = \frac{1}{5}(360 - 11.4 \cdot 15) = 106.2.$$

Den skattade regressionslinjen är alltså

$$y = 106.2 - 11.4x.$$

För konfidensintervallet behövs

$$s^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{1}{3}\left(1320 - \frac{114^2}{10}\right) = 6.8.$$

Då får vi som symmetriskt 95% konfidensintervall

$$b \in \hat{b} \pm F_{t_3}^{-1}(0.975)\frac{s}{\sqrt{S_{xx}}} = -11.4 \pm 3.182\sqrt{\frac{6.8}{10}} \approx -11.4 \pm 2.6.$$

Prediktionsintervallet for en observation Y av prediktionsgrad $1 - \alpha$ vid ett givet värde på x i allmän form är

$$Y \in \hat{a} + \hat{b}x \pm F_{t_{n-2}}^{-1}(1 - \alpha/2)s\sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}}.$$

I vårt fall får vi

$$Y \in 106.2 - 11.4 \cdot 6 \pm 2.353 \cdot \sqrt{6.8}\sqrt{\frac{6}{5} + \frac{3^2}{10}} = 37.8 \pm 8.9$$

med prediktionsgrad 90%.

4. Ett givet flygplan tar 100 passagerare.

(a) (3p) Man vet av erfarenhet att i genomsnitt 5% av dem som köpt biljetter uteblir från sin flight. Därför kan flygbolaget sälja en aning mer än 100 biljetter. Hur många biljetter kan flygbolaget sälja om man vill vara minst 95% säker på inte mer än 100 passagerare ska komma till flighten?

(b) (3p) Om vi för enkelhets skull antar att det kommer exakt 100 passagerare till flighten varav exakt hälften kvinnor och hälften män, vad är sannolikheten att passagerarnas samlade vikt överstiger 7.5 ton? En på måfå vald kvinna kan antas ha en vikt som är normalfördelad med väntevärde 63 och varians 10^2 och detsamma för en på måfå vald man om siffrorna byts till väntevärde 83 och varians 15^2.

Lösning: Del (a): Låt X vara antalet passagerare som uteblir om man bokar n passagerare. Det gäller att $X \sim Bin(n, 0.05)$. Vi vill finna det största n sådant att $\mathbb{P}(X \geq n - 100) \geq 0.95$. Vi får pröva oss fram. Det gäller att om $n = 102$ så är $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - 0.95^{102} - 102 \cdot 0.95^{101} \cdot 0.05 \approx 0.966$ så det går fint att boka 102 passagerare, medan det för $n = 103$ räcker att konstatera att $\mathbb{P}(X = 2) = \binom{103}{2} 0.05^2 \cdot 0.95^{101} \approx 0.074$ för att se att det inte går bra att boka 103 passagerare. Svaret är alltså att vi kan boka max 102 passagerare för att vara minst 95% säkra på att planet inte blir överfullt.

Del (b): Låt $X_1, \ldots, X_{50}$ vara de individuella vikterna hos kvinnorna och $Y_1, \ldots, Y_{50}$ de individulla vikterna för männen. Det är rimligt att anta alla dessa stokastiska variabler är oberoende. Låt $X = \sum_1^{50} X_k$ och $Y = \sum_1^{50} Y_k$. Då gäller att

$$X \sim N(50 \cdot 63, 50 \cdot 10^2) = N(3150, 5000)$$

och

$$Y \sim N(50 \cdot 83, 50 \cdot 15^2) = N(4150, 11250).$$

Den totala vikten av alla passagerare är $T = X + Y$ som blir $N(7300, 16250)$-fördelad. Alltså

$$\mathbb{P}(T \geq 7500) = 1 - \mathbb{P}\left(\frac{T - 7300}{\sqrt{16250}} < \frac{7500 - 7300}{\sqrt{16250}}\right) \approx 1 - \Phi(1.57) \approx 0.058.$$

5. (5p) Betrakta täthetsfunktionen

$$f(x) = \theta^2 x e^{-\theta x},\ x > 0.$$

Här är $\theta > 0$ en okänd parameter. Ett stickprov $X_1, X_2, \ldots, X_n$ på denna fördelning samlas in. Vad blir ML-skattningen av θ?

Lösning: Likelihoodfunktionen för stickprovet är

$$L(\theta; x_1, \ldots, x_n) = \theta^{2n} \left(\prod_{k=1}^{n} x_k\right) e^{-\theta \sum_{k=1}^{n} x_k}.$$

Logaritmera och få

$$l(\theta; x_1, \ldots, x_n) = 2n \ln \theta + \sum_{k=1}^{n} \ln x_k - \theta \sum_{k=1}^{n} x_k.$$

Vi löser $\frac{\partial l}{\partial \theta} = 0$, dvs

$$\frac{2n}{\theta} - \sum_{k=1}^{n} x_k = 0.$$

Lösningen är $\theta = 2n / \sum_{k=1}^{n} x_k = 2/\overline{x}$. ML-skattningen ges alltså av

$$\hat{\theta} = \frac{2}{\overline{x}}.$$

6. (5p) Jensens olikhet:

(a) (3p) Låt $f(x)$ vara en funktion som är två gånger deriverbar med $f''(x) > 0$ för alla x. Visa att det för en stokastisk variabel sådan att både $\mathbb{E}[X]$ och $E[f(X)]$ existerar, gäller att
$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

(b) (2p) Använd resultatet i (a) till att visa att $\ln \mathbb{E}[e^X] \geq \mathbb{E}[X]$ (förutsatt att $\mathbb{E}[e^X]$ existerar). Glöm inte att du kan göra denna del även om du inte löser (a).

Tips för (a): Skriv $\mu = \mathbb{E}[X]$ och observera att tangenten till f i μ ligger under kurvan $f(x)$.

Lösning: Skriv $t(x)$ för tangenten till f i μ. Denna har ekvation $t(x) = f(\mu)+(x-\mu)f'(\mu)$. Eftersom f har positiv andraderivata är f' växande vilket medför att $f'(x) > t'(x)$ för $x > \mu$ och $f'(x) < t'(x)$ för $x < \mu$. Eftersom $f(\mu) = t(\mu)$ medför detta att $f(x) \geq t(x)$ för alla x. Detta medför i sin tur att

$$\mathbb{E}[f(X)] \geq \mathbb{E}[t(X)] = f(\mu) + \mathbb{E}[(X-\mu)]f'(\mu) = f(\mu).$$

För (b) noterar vi att funktionen $f(x) = e^x$ har positiv andraderivata, vilket enligt (a) medför att $\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]}$. Därför får vi

$$\ln \mathbb{E}[e^X] \geq \ln e^{\mathbb{E}[X]} = \mathbb{E}[X].$$

7. Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov på en normalfördelning med okänt väntevärde μ och okänd varians σ^2.

(a) (2p) Gör ett nedåt begränsat 95% konfidensintervall för σ^2.

(b) (2p) Hur stort måste n vara för att styrkan i testet $H_0 : \sigma^2 = 1$ mot $H_A : \sigma^2 > 1$ på 5% signifikansnivå ska vara minst 95% för $\sigma^2 = 4$?

Lösning: Utnyttja att $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$. Detta betyder att

$$\mathbb{P}\left(\frac{(n-1)s^2}{\sigma^2} \leq F^{-1}_{\chi^2_{n-1}}(0.95)\right) = 0.95.$$

Detta ger det nedåt begränsade konfidensintervallet

$$\sigma^2 \geq \frac{(n-1)s^2}{F^{-1}_{\chi^2_{n-1}}(0.95)}.$$

Detta ger oss ett test av $H_0 : \sigma^2 = 1$ mot $H_A : \sigma^2 > 1$ som på 5% signifikansnivå förkastar H_0 om

$$(n-1)s^2 \geq F^{-1}_{\chi^2_{n-1}}(0.05).$$

Om $\sigma^2 = 4$ är det korrekta värdet så gäller att $(n-1)s^2/4 \sim \chi^2_{n-1}$ och eftersom testet förkastar om $(n-1)s^2/4 \geq F^{-1}_{\chi^2_{n-1}}(0.95)/4$ gäller att sannolikheten att förkasta är

$$1 - F_{\chi^2_{n-1}}\left(\frac{F^{-1}_{\chi^2_{n-1}}(0.95)}{4}\right).$$

Denna sannolikhet överstiger 0.95 då

$$F^{-1}_{\chi^2_{n-1}}(0.95) \leq 4F^{-1}_{\chi^2_{n-1}}(0.05).$$

Enligt tabell gäller detta för $n-1 \geq 13$, men inte för $n-1 = 12$. Alltså krävs $n-1 \geq 13$, dvs $n \geq 14$.

8. (3p) Låt X vara en stokastisk variabel som kan anta värdena 0, 1 eller 2 och låt M_X vara dess momentgenererande funktion. Bestäm p_k, $k = 0, 1, 2$, då $M_X(1) = 1/6 + e/2 + e^2/3$ och $M_X(2) = 1/6 + e^2/2 + e^4/3$.

Lösning: Per definition är $M_X(1) = \mathbb{E}[e^X] = p_0 + p_1 e + p_2 e^2$ och $M_X(2) = p_0 + p_1 e^2 + p_2 e^4$ och eftersom vi också har $p_1 + p_2 + p_3 = 1$ ger detta tre linjärt oberoende linjära ekvationer för p_0, p_1 och p_2 varför detta ekvationssystem har en unik lösning. En sådan lösning kan vi läsa ut direkt ur uttrycken för $M_X(1)$ och $M_X(2)$, nämligen $p_0 = 1/6$, $p_1 = 1/2$, $p_2 = 1/3$.

Lycka till!
Johan Jonasson

Tabell 1: Values of the cdf $\Phi(x)$ of the standard normal distribution [e.g., $\Phi(1.41) = 0.921$]

x	0	1	2	3	4	5	6	7	8	9
0.0	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1	.540	.544	.548	.552	.556	.560	.564	.568	.571	.575
0.2	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5	.692	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6	.726	.729	.732	.736	.739	.742	.745	.749	.752	.755
0.7	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813
0.9	.816	.819	.821	.824	.826	.829	.832	.834	.836	.839
1.0	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.892	.894	.896	.898	.900	.902
1.3	.903	.905	.907	.908	.910	.912	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932
1.5	.933	.934	.936	.937	.938	.939	.941	.942	.943	.944
1.6	.945	.946	.947	.948	.950	.951	.952	.952	.9545	.954
1.7	.955	.956	.957	.958	.959	.960	.961	.962	.962	.963
1.8	.964	.965	.966	.966	.967	.968	.969	.969	.970	.971
1.9	.971	.972	.973	.973	.974	.974	.975	.976	.976	.977
2.0	.977	.978	.978	.979	.979	.980	.980	.981	.981	.982
2.1	.982	.983	.983	.983	.984	.984	.985	.985	.985	.986
2.2	.986	.986	.987	.987	.988	.988	.988	.988	.989	.989
2.3	.989	.990	.990	.990	.990	.991	.991	.991	.991	.992
2.4	.992	.992	.992	.992	.993	.993	.993	.993	.993	.994
2.5	.994	.994	.994	.994	.995	.995	.995	.995	.995	.995
2.6	.995	.996	.996	.996	.996	.996	.996	.996	.996	.996
2.7	.996	.997	.997	.997	.997	.997	.997	.997	.997	.997
2.8	.997	.998	.998	.998	.998	.998	.998	.998	.998	.998
2.9	.998	.998	.998	.998	.998	.998	.998	.998	.999	.999

Tabell 2: Values of $\Phi(x)$ commonly used in confidence intervals and tests, and the corresponding x values

$\Phi(x)$	0.90	0.95	0.975	0.99	0.995
x	1.28	1.64	1.96	2.33	2.58

Tabell 3: Percentiles of the t distribution with DF degrees of freedom [e.g., $F_{t_7}(1.89) = 0.95$]

DF	0.95	0.975	0.99	0.995	DF	0.95	0.975	0.99	0.995
1	6.31	12.71	31.82	63.66	16	1.75	2.12	2.58	2.92
2	2.92	4.30	6.96	9.92	17	1.74	2.11	2.58	2.90
3	2.35	3.18	4,54	5.84	18	1.73	2.10	2.55	2.88
4	2.13	2.78	3.74	4.60	19	1.73	2.09	2.54	2.86
5	2.02	2.57	3.36	4.03	20	1.72	2.09	2.53	2.85
6	1.94	2.45	3.14	3.71	21	1.72	2.08	2.52	2.83
7	1.89	2.36	3.00	3.50	22	1.72	2.07	2.51	2.82
8	1.86	2.31	2.90	3.36	23	1.71	2.07	2.50	2.81
9	1.83	2.26	2.82	3.25	24	1.71	2.06	2.49	2.80
10	1.81	2.23	2.76	3.17	25	1.71	2.06	2.49	2.79
11	1.80	2.20	2.72	3.11	26	1.71	2.06	2.48	2.78
12	1.78	2.18	2.68	3.05	27	1.70	2.05	2.47	2.77
13	1.77	2.16	2.65	3.01	28	1.70	2.05	2.47	2.76
14	1.76	2.14	2.62	2.98	29	1.70	2.05	2.46	2.76
15	1.75	2.13	2.60	2.95	30	1.70	2.04	2.46	2.75

Tabell 4: Percentiles of the chi-square distribution with DF degrees of freedom [e.g., $F_{\chi^2_{20}}(10.85) = 0.05$]

DF	0.025	0.05	0.95	0.975	DF	0.025	0.05	0.95	0.975
1	0.001	0.004	3.84	5.02	16	6.91	7.96	26.30	28.84
2	0.05	0.10	5.99	7.38	17	7.56	8.67	27.59	30.19
3	0.22	0.35	7.82	9.34	18	8.23	9.39	28.87	31.53
4	0.48	0.71	9.49	11.14	19	8.91	10.12	30.14	32.85
5	0.83	1.14	11.07	12.83	20	9.59	10.85	31.41	34.17
6	1.24	1.64	12.59	14.45	21	10.28	11.60	32.67	35.48
7	1.69	2.17	14.07	16.01	22	10.98	12.34	33.92	36.78
8	2.18	2.73	15.51	17.54	23	11.69	13.09	35.17	38.08
9	2.70	3.32	19.92	19.02	24	12.40	13.85	36.42	39.36
10	3.25	3.94	18.31	20.48	25	13.12	14.61	37.65	40.65
11	3.82	4.58	19.68	21.92	26	13.84	15.38	38.88	41.92
12	4.40	5.23	21.03	23.34	27	14.57	16.15	40.11	43.19
13	5.01	5.89	22.36	27.74	28	15.31	16.93	41.34	44.46
14	5.63	6.57	23.68	26.12	29	16.05	17.71	42.56	45.72
15	6.26	7.26	25.00	27.49	30	16.79	18.49	43.77	46.98

Tabell 5: Percentiles of the F distribution with r and s degrees of freedom [e.g., $F_{F_{8,20}}(2.45) = 0.95$]

	2.5 % percentile								
s	$r = 2$	3	4	5	6	7	8	9	10
2	0.026	0.062	0.094	0.119	0.138	0.153	0.165	0.175	0.183
3	0.026	0.065	0.100	0.129	0.152	0.170	0.185	0.197	0.207
4	0.025	0.066	0.104	0.135	0.161	0.181	0.198	0.212	0.224
5	0.025	0.067	0.107	0.140	0.167	0.189	0.208	0.223	0.236
6	0.025	0.068	0.109	0.143	0.172	0.195	0.215	0.231	0.246
7	0.025	0.068	0.110	0.146	0.176	0.200	0.221	0.238	0.253
8	0.025	0.069	0.111	0.148	0.179	0.204	0.226	0.244	0.259
9	0.025	0.069	0.112	0.150	0.181	0.207	0.230	0.248	0.265
10	0.025	0.069	0.113	0.151	0.183	0.210	0.233	0.252	0.269
12	0.025	0.070	0.114	0.153	0.186	0.214	0.238	0.259	0.276
15	0.025	0.070	0.116	0.156	0.190	0.219	0.244	0.265	0.284
16	0.025	0.070	0.116	0.156	0.191	0.220	0.245	0.267	0.286
18	0.025	0.070	0.116	0.157	0.192	0.222	0.248	0.270	0.290
20	0.025	0.071	0.117	0.158	0.193	0.224	0.250	0.273	0.293
21	0.025	0.071	0.117	0.158	0.194	0.225	0.251	0.274	0.294
24	0.025	0.071	0.117	0.159	0.195	0.227	0.253	0.277	0.297
25	0.025	0.071	0.118	0.160	0.196	0.227	0.254	0.278	0.298
27	0.025	0.071	0.118	0.160	0.197	0.228	0.255	0.279	0.300
28	0.025	0.071	0.118	0.160	0.197	0.228	0.256	0.280	0.301
30	0.025	0.071	0.118	0.161	0.197	0.229	0.257	0.281	0.302

	95 % percentile								
s	$r = 2$	3	4	5	6	7	8	9	10
2	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
12	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
15	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
18	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
20	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
24	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
27	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
30	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16

	97.5 % percentile								
s	$r=2$	3	4	5	6	7	8	9	10
2	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
3	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
4	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
5	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
6	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
7	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
8	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30
9	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
10	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
12	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37
15	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
16	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99
18	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87
20	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77
21	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73
24	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64
25	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61
27	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57
28	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55
30	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51

Tabell 6: Critical values c for the Wilcoxon signed rank test, where n is the sample size and $C = n(n+1) - c$ [e.g., if $n = 20$, then $P(W \leq 61) = P(W \geq 149) \approx 0.05$]

n	0.025	0.05	$n(n+1)/2$	n	0.025	0.05	$n(n+1)/2$
5	0	1	15	18	41	48	171
6	1	3	21	19	47	54	190
7	3	4	28	20	53	61	210
8	4	6	36	21	59	68	231
9	6	9	45	22	67	76	253
10	9	11	55	23	74	84	276
11	11	14	66	24	82	92	300
12	14	18	78	25	90	101	325
13	18	22	91	26	99	111	351
14	22	26	105	27	108	120	378
15	26	31	120	28	117	131	406
16	30	36	136	29	127	141	435
17	35	42	153	30	138	152	465

Tabell 7: Critical values c for the Wilcoxon rank sum test, where m is the size of the smaller sample, and $C = m(m+n+1) - c$ [e.g., if $m = 4$ and $n = 8$, then $P(W \leq 16) = P(W \geq 36) \approx 0.05$]

n	$P(W \leq c)$	$m = 2$	3	4	5	6	7	8	9	10	11
2	0.025	3									
	0.05	3									
3	0.025	3	3								
	0.05	6	7								
4	0.025	3	6	11							
	0.05	3	7	12							
5	0.025	3	7	12	18						
	0.05	4	8	13	20						
6	0.025	3	8	13	19	27					
	0.05	4	9	14	21	29					
7	0.025	3	8	14	21	28	37				
	0.05	4	9	15	22	30	40				
8	0.025	4	9	15	22	30	39	50			
	0.05	5	10	16	24	32	42	52			
9	0.025	4	9	15	23	32	41	52	63		
	0.05	5	11	17	25	34	44	55	67		
10	0.025	4	10	16	24	33	43	54	66	79	
	0.05	5	11	18	27	36	46	57	70	83	
11	0.025	5	10	17	25	35	45	56	69	82	97
	0.05	5	12	19	28	38	48	60	73	87	101