Tentamen
MVE302 Sannolikhet och statistik

2021-10-09 kl. 14.00 - 18.00

Examinator: Johan Jonasson, Matematiska vetenskaper, Chalmers

Telefonvakt: Johan Jonasson, telefon: 031-7723546

Hjälpmedel: Valfri miniräknare och fyra A4-sidor (två blad) med handskrivna anteckningar.

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. (5p) En urna innehåller tre blå, tre röda och fyra gröna bollar. Tre av urnans tio bollar väljs på måfå. Låt X vara antalet blå bollar bland de valda och låt Y vara antalet röda bollar bland de valda bollarna. Beräkna frekvensfunktion, väntevärde och varians för XY. Beräkna också $\mathbb{C}\text{ov}(X, Y)$.

2. (5p) Låt X vara en diskret stokastisk variabel med $p_X(1) = 1/2$, $p_X(2) = 1/3$ och $p_X(3) = 1/6$. Sedan, givet $X = x$, låt $Y(t)$ vara en Poissonprocess med intensitet x.

(a) Vad är frekvensfunktionen för $Y(t)$?

(b) Vad är $\mathbb{P}(X = x|Y(2) = 3)$, $x = 1, 2, 3$.

3. Låt $X_1, \ldots, X_n$ vara ett stickprov på en kontinuerlig stokastisk variabel som har täthetsfunktion

$$f_\theta(x) = C(\theta)x^{\theta-1},\ x \in [0, 1],$$

där θ är en okänd positiv parameter. Bestäm $C(\theta)$ och gör en ML-skattning av θ.

4. (5p) Betrakta den vanliga linjära regressionsmodellen, dvs $(x_1, Y_1), \ldots, (x_n, Y_n)$ är sådana att

$$Y_k = a + bx_k + \epsilon_k$$

där ϵ_k:na är oberoende och normalfördelade med väntevärde 0 och varians σ^2 och a, b och σ^2 är okända parametrar. Antag nu att man observerat data $(0, 1)$, $(2, 4)$ och $(5, 12)$. Skatta a och b och ge ett 95% konfidensintervall för b. Ge också ett 90% prediktionsintervall för Y, där (x, Y) är en ny observation med $x = 3$.

5. (5p) Låt $X = (X_1, \ldots, X_7)$ och $Y = (Y_1, \ldots, Y_6)$ vara två oberoende stickprov på varsin normalfördelning med parametrar μ_x och σ_x^2 respektive μ_y och σ_y^2. Gör ett symmetriskt 95% konfidensintervall för $\mu_x - \mu_y$ i följande två fall

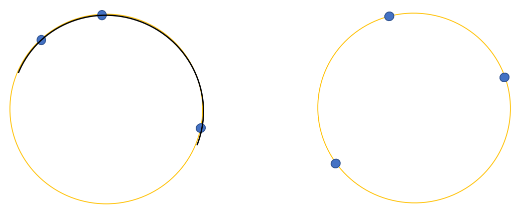
(a) Du känner till att $\sigma_x^2 = 3^2$ och $\sigma_y^2 = 4^2$.

(b) Du känner inte till varken σ_x^2 eller σ_y^2, men det verkar ok att anta att $\sigma_x = \sigma_y$.

Data var $X = (3.1,\ 2.9,\ 5.6,\ 9.2,\ 5.9,\ 2.0,\ 3.9)$ och $Y = (-0.5,\ 0.0,\ -0.9,\ 5.1,\ 4.2,\ 0.7)$.

6. (5p)

(a) Placera ut tre punkter på måfå på en cirkel. Vad är sannolikheten att punkterna placerar sig så att de kan täckas av en halvcirkel? (Med andra ord, vad är sannolikheten någon av punkterna hamnar på den kortare av de två sträckorna längs cirkeln mellan de två andra punkterna?) Se figur.



Figur 1: Till vänster tre punkter som kan täckas av en halvcirkel och till höger tre punkter som inte kan täckas av en halvcirkel.

(b) Placera ut två punkter på måfå på $[-1,1] \times [-1,1]$. Vad är sannolikheten att dessa två punkter kan täckas av en kvadrat med sidan 1?

7. (5p) Låt $X_1, \ldots, X_n$ vara ett stickprov på en normalfördelning med okänt väntevärde μ och känd varians $\sigma^2 = 1$.

- Om det korrekta värdet på μ är 2, hur stort måste n vara för styrkan i testet $H_0 : \mu = 0$ mot $H_A : \mu > 0$ på 95% signifikansnivå ska vara 0.9?
- Om $n = 10$, vad är sannolikheten att ett symmetriskt 95% konfidensintervall för μ blir kortare för den som inte känner till σ^2 än för den som gör det?

8. (5p) Antag att parametern Θ har en priorfördelning som är $\Gamma(2, \beta)$. Observera ett stickprov $X_1, \ldots, X_n$ på en Poissonfördelning med parameter Θ. Vad blir posteriorfördelningen för Θ givet stickprovet?

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

Tentamen
MVE302 Sannolikhet och statistik

2021-10-09 kl. 14.00 - 18.00

Examinator: Johan Jonasson, Matematiska vetenskaper, Chalmers

Telefonvakt: Johan Jonasson, telefon: 031-7723546

Hjälpmedel: Valfri miniräknare och fyra A4-sidor (två blad) med handskrivna anteckningar.

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. (5p) En urna innehåller tre blå, tre röda och fyra gröna bollar. Tre av urnans tio bollar väljs på måfå. Låt X vara antalet blå bollar bland de valda och låt Y vara antalet röda bollar bland de valda bollarna. Beräkna frekvensfunktion, väntevärde och varians för XY. Beräkna också $\mathbb{C}\text{ov}(X, Y)$.

Lösning. XY kan anta värdena $0, 1, 2$. Antalet sätt att väja tre bollar av tio är $\binom{10}{3}$. Antalet sätt att välja två blå och en röd boll eller vice versa är $2 \cdot 3 \cdot 3$. Antalet sätt att välja en blå, en grön och en röd är $3 \cdot 3 \cdot 4$. Alltså

$$\mathbb{P}(XY = 2) = \frac{2 \cdot 3 \cdot 3}{\binom{10}{3}} = \frac{3}{20}.$$

$$\mathbb{P}(XY = 1) = \frac{3 \cdot 3 \cdot 4}{\binom{10}{3}} = \frac{3}{10}.$$

Därmed blir

$$\mathbb{P}(XY = 0) = 1 - \frac{3}{20} - \frac{3}{10} = \frac{11}{20}.$$

Nu får vi

$$\mathbb{E}[XY] = 1 \cdot \frac{3}{10} + 2 \cdot \frac{3}{20} = \frac{3}{5}$$

och

$$\mathbb{E}[(XY)^2] = 1 \cdot \frac{3}{10} + 4 \cdot \frac{3}{20} = \frac{9}{10}.$$

Detta ger

$$\mathbb{V}\text{ar}(XY) = \frac{9}{10} - \frac{3^2}{5^2} = \frac{27}{50}.$$

Det är uppenbart att $\mathbb{E}[X] = \mathbb{E}[Y] = 9/10$, så till sist får vi

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{3}{5} - \frac{9^2}{10^2} = -\frac{21}{100}.$$

2. (5p) Låt X vara en diskret stokastisk variabel med $p_X(1) = 1/2$, $p_X(2) = 1/3$ och $p_X(3) = 1/6$. Sedan, givet $X = x$, låt $Y(t)$ vara en Poissonprocess med intensitet x.

(a) Vad är frekvensfunktionen för $Y(t)$?

(b) Vad är $\mathbb{P}(X = x|Y(2) = 3)$, $x = 1, 2, 3$.

Lösning. Det gäller att
$$\mathbb{P}(Y(t) = k|X = x) = e^{-xt}\frac{(xt)^k}{k!}.$$
Enligt TSL blir nu
$$\mathbb{P}(Y(t) = k) = \frac{1}{2}e^{-t}\frac{t^k}{k!} + \frac{1}{3}e^{-2t}\frac{(2t)^k}{k!} + \frac{1}{6}e^{-3t}\frac{(3t)^k}{k!}.$$
Enligt Bayes formel är
$$\begin{aligned}\mathbb{P}(X = x|Y(2) = 3) &= \frac{\mathbb{P}(Y(2) = 3|X = x)\mathbb{P}(X = x)}{\mathbb{P}(Y(2) = 3)} \\ &= \frac{e^{-2x}\frac{(2x)^3}{3!}\mathbb{P}(X = x)}{\frac{1}{2}e^{-2}\frac{2^3}{3!} + \frac{1}{3}e^{-4}\frac{4^3}{3!} + \frac{1}{6}e^{-6}\frac{6^3}{3!}} \\ &= \frac{e^{-2x}(2x)^3\mathbb{P}(X = x)}{4e^{-2} + \frac{64}{3}e^{-4} + 36e^{-6}}.\end{aligned}$$

3. Låt $X_1, \ldots, X_n$ vara ett stickprov på en kontinuerlig stokastisk variabel som har täthetsfunktion
$$f_\theta(x) = C(\theta)x^{\theta-1},\ x \in [0, 1],$$
där θ är en okänd positiv parameter. Bestäm $C(\theta)$ och gör en ML-skattning av θ.

Lösning. Storheten $C(\theta)$ bestäms av att
$$C(\theta)\int_0^1 x^{\theta-1}dx = 1,$$
vilket ger $C(\theta) = \theta$. Detta ger sedan likelihood
$$L(\theta; x_1, \ldots, x_n) = \theta^n \prod_k x_k^\theta$$
som efter logaritmering ger
$$\ell(\theta; x_1, \ldots, x_n) = n \ln\theta + (\theta - 1)\sum_k \ln x_k.$$
Derivera och sätt till 0 och lös ut θ. Detta ger
$$\hat{\theta} = -\frac{n}{\sum_{k=1}^n \ln x_k}.$$

4. (5p) Betrakta den vanliga linjära regressionsmodellen, dvs $(x_1, Y_1), \ldots, (x_n, Y_n)$ är sådana att
$$Y_k = a + bx_k + \epsilon_k$$
där ϵ_k:na är oberoende och normalfördelade med väntevärde 0 och varians σ^2 och a, b och σ^2 är okända parametrar. Antag nu att man observerat data $(0, 1)$, $(2, 4)$ och $(5, 12)$. Skatta a och b och ge ett 95% konfidensintervall för b. Ge också ett 90% prediktionsintervall för Y, där (x, Y) är en ny observation med $x = 3$.

Lösning. Det gäller att $\hat{b} = S_{xy}/S_{xx} = 85/38 \approx 2.24$ och $\hat{a} = \overline{y} - \hat{b}\overline{x} = 17/38 \approx 0.45$. Vidare är
$$s^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{49}{38}.$$

Därmed

$$b = \hat{b} \pm F_{t_1}^{-1}(0.975)\frac{s}{S_{xx}} = \frac{85}{38} \pm 12.7\sqrt{\frac{49/38}{38/3}} \approx 2.24 \pm 4.06 \ (95\%).$$

Prediktionsintervallet är

$$Y = \hat{a} + \hat{b}x \pm F_{t_1}^{-1}(0.95)s\sqrt{\frac{4}{3} + \frac{(x-\overline{x})^2}{S_{xx}}} \approx 7.13 \pm 8.38 \ (90\%).$$

5. (5p) Låt $X = (X_1, \ldots, X_7)$ och $Y = (Y_1, \ldots, Y_6)$ vara två oberoende stickprov på varsin normalfördelning med parametrar μ_x och σ_x^2 respektive μ_y och σ_y^2. Gör ett symmetriskt 95% konfidensintervall för $\mu_x - \mu_y$ i följande två fall

(a) Du känner till att $\sigma_x^2 = 3^2$ och $\sigma_y^2 = 4^2$.

(b) Du känner inte till varken σ_x^2 eller σ_y^2, men det verkar ok att anta att $\sigma_x = \sigma_y$.

Data var $X = (3.1, 2.9, 5.6, 9.2, 5.9, 2.0, 3.9)$ och $Y = (-0.5, 0.0, -0.9, 5.1, 4.2, 0.7)$.

Lösning. I del (a) utnyttjar vi att

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{7} + \frac{\sigma_y^2}{6}}} \sim N(0,1).$$

Detta ger

$$\begin{aligned}
\mu_x - \mu_y &= \overline{X} - \overline{Y} \pm \Phi^{-1}(0.975)\sqrt{\frac{\sigma_x^2}{7} + \frac{\sigma_y^2}{6}} \\
&= 4.7 - 1.4 \pm 1.96\sqrt{3^2/7 + 4^2/6} \\
&= 3.3 \pm 3.9.
\end{aligned}$$

I (b) får vi utnyttja att

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{s\sqrt{\frac{1}{7} + \frac{1}{6}}} \sim t_{11}.$$

Här är

$$s^2 = \frac{6s_x^2 + 5s_y^2}{11} = 6.57.$$

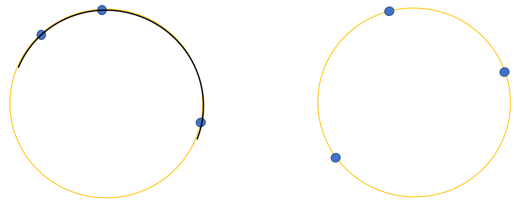
Konfidensintervallet blir

$$\begin{aligned}
\mu_x - \mu_y &= \overline{X} - \overline{Y} \pm F_{t_{11}}^{-1}(0.975)s\sqrt{\frac{1}{7} + \frac{1}{6}} \\
&= 3.3 \pm 2.20 \cdot \sqrt{6.57} \cdot 0.56 \\
&= 3.3 \pm 3.1.
\end{aligned}$$

6. (5p)

(a) Placera ut tre punkter på måfå på en cirkel. Vad är sannolikheten att punkterna placerar sig så att de kan täckas av en halvcirkel? (Med andra ord, vad är sannolikheten någon av punkterna hamnar på den kortare av de två sträckorna längs cirkeln mellan de två andra punkterna?) Se figur.

(b) Placera ut två punkter på måfå på $[-1,1] \times [-1,1]$. Vad är sannolikheten att dessa två punkter kan täckas av en kvadrat med sidan 1?



Figur 1: Till vänster tre punkter som kan täckas av en halvcirkel och till höger tre punkter som inte kan täckas av en halvcirkel.

Lösning. Låt X, Y och Z vara positionerna för de tre punkterna, räknade som avståndet medurs från cirkelns nordpol (kl 12). Det är uppenbart att omkretsen av cirkeln inte har någon betydelse, så antag att omkretsen är 2. Vi kan också uppenbarligen anta att $Z = 0$ (dvs att motsvarande punkt ligger på nordpolen). Ytterligare ett uppenbart faktum är att vi pga symmetri kan anta att $0 < X < 1$, vilket ger $f_X(x) = 1$. Vi har också $f_Y(y) = 1/2$, $0 < y < 2$. Använd nu TSL genom att betinga på X. Vi kallar händelsen vars sannolikhet eftersöks för H. Observera att om $X = x$, så inträffar H om Y ligger mellan punkten på avstånd $1-x$ moturs från nordpolen och punkten $1-x$ medurs från x, ett intervall av längd $2-x$, och sannolikheten att Y hamnar där är $(2-x)/2$. Med andra ord $\mathbb{P}(H|X = x) = (2-x)/2$. Detta ger

$$\mathbb{P}(H) = \int_0^1 \mathbb{P}(H|X = x) f_X(x) dx = \int_0^1 \frac{2-x}{2} dx = \frac{3}{4}.$$

Del (b) är väldigt lik del (a). Kalla de två punkterna som placeras ut för $X = (X_1, X_2)$ och $Y = (Y_1, Y_2)$. Det gäller att X_1, X_2, Y_1, Y_2 är oberoende och likformigt fördelade på $[-1, 1]$. Dock ser vi att av symmetri kan vi anta att X_1 och X_2 är likformigt fördelade på $[0, 1]$. Om $(X_1, X_2) = (x_1, x_2)$, så inser vi efter en stunds eftertanke att H inträffar då $Y_1 > x_1 - 1$ och $Y_2 > x_2 - 1$. Detta ger nu $\mathbb{P}(H|X_1 = x_1, X_2 = x_2) = \frac{(2-x_1)(2-x_2)}{4}$. TSL ger (eftersom tätheten för (X_1, X_2) är 1)

$$\mathbb{P}(H) = \frac{1}{4} \int_0^1 \int_0^1 (2-x_1)(2-x_2) dx_1 dx_2 = \frac{9}{16}.$$

7. (5p) Låt $X_1, \ldots, X_n$ vara ett stickprov på en normalfördelning med okänt väntevärde μ och känd varians $\sigma^2 = 1$.

- Om det korrekta värdet på μ är 2, hur stort måste n vara för styrkan i testet $H_0 : \mu = 0$ mot $H_A : \mu > 0$ på 95% signifikansnivå ska vara 0.9?
- Om $n = 10$, vad är sannolikheten att ett symmetriskt 95% konfidensintervall för μ blir kortare för den som inte känner till σ^2 än för den som gör det?

Lösning. Del (a): Testet förkastar H_0 om $\sqrt{n}\bar{X} \geq 1.64$. Om $\mu = 2$ gäller

$$\begin{aligned}\mathbb{P}(\sqrt{n}\bar{X} \geq 1.64) &= \mathbb{P}(\sqrt{n}(\bar{X} - 2) \geq 1.64 - 2\sqrt{n}) \\ &= \Phi(2\sqrt{n} - 1.64).\end{aligned}$$

Detta blir minst 0.9 då $2\sqrt{n} - 1.64 > 1.28$, vilket ger $n > 2.13$. Svaret är då att det krävs att $n \geq 3$ för att nå den önskade styrkan.

Längden av konfidensintervallet för den som känner till σ^2 blir $1.96\sigma/\sqrt{10}$ medan den för den som inte känner till σ^2 blir $F_{t_9}^{-1}(0.975)s/\sqrt{10} = 2.26s/\sqrt{10}$. Det senare blir kortare än det första om

$$\frac{s^2}{\sigma^2} < \frac{1.96^2}{2.26^2} = 0.75.$$

Nu utnyttjar vi att $9s^2/\sigma^2 \sim \chi^2_9$, vilket ger att den sökta sannolikheten är

$$F_{\chi^2_9}(9 \cdot 0.75) = 0.10.$$

8. (5p) Antag att parametern Θ har en priorfördelning som är $\Gamma(2, \beta)$. Observera ett stickprov $X_1, \ldots, X_n$ på en Poissonfördelning med parameter Θ. Vad blir posteriorfördelningen för Θ givet stickprovet?

Lösning. Enligt Bayes formel är

$$f_{\Theta|X}(\theta|x) = \frac{f_\Theta(\theta) f_{X|\Theta}(x|\theta)}{f_X(x)}$$

där $X = (X_1, \ldots, X_n)$ och $x = (x_1, \ldots, x_n)$ är de observerade värdena. I det följande bortses från konstanter som inte beror av θ eftersom priorfördelningen för Θ är en funktion av θ. Då erhålls

$$\begin{aligned} f_{\Theta|X}(\theta|x) &\propto f_\Theta(\theta) f_{X|\Theta}(x|\theta) \\ &\propto \theta e^{-\beta\theta} \prod_k \left(e^{-\theta} \frac{\theta^{x_k}}{x_k!} \right) \\ &\propto \theta^{1+\sum_k x_k} e^{-(\beta+n)\theta}. \end{aligned}$$

Detta känner vi igen som en ny gammafördelning, nämligen $\Gamma(2 + \sum_{k=1}^n x_k, \beta + n)$, vilket alltså är posteriorfördelningen.

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