Tuesday 14th, May 2019
2.00 p.m. - 4.00 p.m.

EXAMINATION FOR THE DEGREE OF BSc, MSci and MSc (Taught) (SCIENCE)

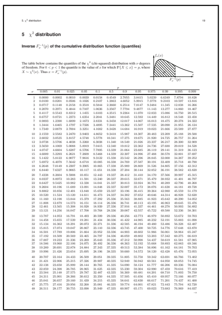
Statistics – Generalised Linear Models - Level M

This paper consists of 9 pages and contains 4 questions.
Candidates should attempt all questions.

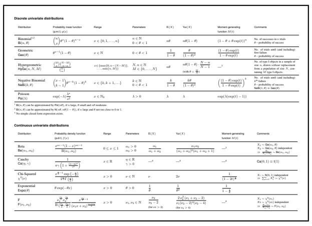
Question 1	20 marks
Question 2	20 marks
Question 3	20 marks
Question 4	20 marks
Total	80 marks

The following material is made available to you:

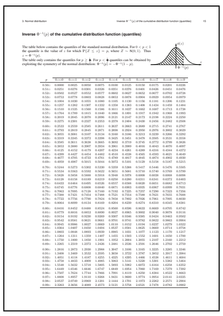
Chisquared Distribution



Probability formula sheet



Normal Distribution



CONTINUED OVERLEAF/

"Hand calculators with simple basic functions (log, exp, square root, etc.) may be used in examinations. No calculator which can store or display text or graphics may be used, and any student found using such will be reported to the Clerk of Senate".

1. (a) Suppose that Y is a random variable with a distribution that is a member of the exponential family with probability density or mass function of the form

$$f(y) = \exp[yb(\theta) + c(\theta) + d(y)].$$

Define the *score function* for a single observation and using its properties or otherwise, show that

$$E(Y) = -\frac{c'(\theta)}{b'(\theta)}, \tag{1}$$

where $b'(\theta)$ and $c'(\theta)$ denote the first derivatives of $b(\theta)$ and $c(\theta)$ with respect to θ. **[4 MARKS]**

(b) Show that the $Po(\theta)$ distribution with probability mass function

$$f(y; \theta) = \frac{\theta^y e^{-\theta}}{y!}, \ \theta > 0, \ y = 0, 1, 2, \ldots,$$

is a member of the above exponential family of distributions and use equation (1) from part (a) to obtain an expression for its mean. **[4 MARKS]**

(c) For each of the two studies below, briefly describe a suitable generalised linear model which assumes a Poisson distribution for the response. Explain the role of the Poisson distribution in each model, state your choice of link function and give a simple structural model that might be investigated.

(i) Tomatoes, broccoli and carrots are three foods thought to contain cancer-fighting components. In a nutritional study, researchers attempt to investigate dietary patterns in a population, to understand the patterns of combined consumption of these components that occur. Subjects in a random sample of the population are surveyed and, based on their dietary histories, classified as high or low consumers of each of the three different foods. The results are described in a $2 \times 2 \times 2$ contingency table. **[6 MARKS]**

CONTINUED OVERLEAF/

(ii) Serious infections acquired during hospitalisation are recorded and counted over a five-year period for various hospitals in the US, classified as teaching or non-teaching, public or private, and religious vs secularly-run hospitals. Only some hospitals were in the study for the full five-year period. Also, the lengths of hospital stays vary considerably, and the hospitals are of different sizes with different numbers of occupied beds each day. For these reasons, it is decided to divide the count in each hospital by the average number of persons/day in the hospital, multiplied by the number of days that the hospital was in the study, before comparisons are made between different types of hospitals.

[6 MARKS]

2. The table below shows results of a bioassay to compare the biological potencies of two preparations (batches) of insulin, by measuring the proportions of rodents that exhibit a particular response to different doses of each preparation.

Obs	Prep	Dose	Resp	Total
1	Standard	3.40	0	33
2	Standard	5.20	5	32
3	Standard	7.00	11	38
4	Standard	8.50	14	37
5	Standard	10.50	18	40
6	Standard	13.00	21	37
7	Standard	18.00	23	31
8	Standard	21.00	30	37
9	Standard	28.00	27	30
10	Test	6.50	2	40
11	Test	10.00	10	30
12	Test	14.00	18	40
13	Test	21.50	21	35
14	Test	29.00	27	37

A different group of rodents is used for each dose of each preparation. It is desired to measure the potency of the test preparation relative to the standard where, for example, the potency is 2 if the proportion of responses produced by the standard preparation can be obtained using only half the dose of the test preparation.

Potency may be estimated by using logistic regression analysis, assuming that the rodents exhibit a logistic tolerance distribution in relationship to the log(dose). In this analysis, $\log_{10}$(dose) is used as one explanatory variable and preparation (test or standard) as another. The following is abbreviated output from fitting two models, `m1` and `m0`, in R.

CONTINUED OVERLEAF/

```
> m1 <- glm(cbind(Resp,NonResp)~ log10(Dose)*Prep, family=binomial)
> summary(m1)

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)           -5.7907     0.6839  -8.467   <2e-16 ***
log10(Dose)            5.5180     0.6446   8.561   <2e-16 ***
PrepTest              -0.2170     1.2077  -0.180    0.857
log10(Dose):PrepTest  -0.6269     1.0464  -0.599    0.549

    Null deviance: 166.8335  on 13  degrees of freedom
Residual deviance:   8.4351  on 10  degrees of freedom
AIC: 64.287

> m0 <-  glm(cbind(Resp,NonResp)~ log10(Dose)+Prep, family=binomial)
> summary(m0)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.5531     0.5427  -10.23  < 2e-16 ***
log10(Dose)   5.2894     0.5057   10.46  < 2e-16 ***
PrepTest     -0.9290     0.2334   -3.98 6.89e-05 ***

    Null deviance: 166.8335  on 13  degrees of freedom
Residual deviance:   8.7912  on 11  degrees of freedom
AIC: 62.644
```

(a) Based on the above output, does the effect of preparation on the odds of the response depend on the dose? Explain. **[2 MARKS]**

(b) Give your choice of model (`m1` or `m0`) for estimating the potency. **[2 MARKS]**

(c) Write down the regression equations corresponding to your chosen model. **[4 MARKS]**

(d) Use the above equations to estimate the median effective **actual** dose (**not** the $\log_{10}$(Dose)) of (i) the test preparation, and (ii) the standard preparation, that would be predicted to produce responses in half the rodents. **[4 MARKS]**

(e) The ratio of the median effective doses of standard to test preparations is an estimate of the potency. Calculate the potency using the results of (d) above. **[2 MARKS]**

(f) What is the odds ratio associated with a doubling of the actual dose of either preparation? **[3 MARKS]**

CONTINUED OVERLEAF/

(g) Comment on the goodness of fit of the model you have used to answer parts (b)-(f). You may use the above R output and/or the fitted values given in the table overleaf.

	Observed values	Fitted values m1	Fitted values m0
1	0	1.79	2.00
2	5	4.39	4.67
3	11	9.30	9.61
4	14	12.59	12.80
5	18	18.44	18.49
6	21	21.76	21.61
7	23	23.46	23.18
8	30	30.28	29.92
9	27	26.99	26.73
10	2	4.64	4.05
11	10	7.40	6.98
12	18	16.03	15.86
13	21	21.86	22.32
14	27	28.07	28.78

[3 MARKS]

CONTINUED OVERLEAF/

3. Data were collected on 44 doctors working in an emergency service at a hospital to study the factors affecting the number of complaints received during a one-year period. In addition to the response, Y_i, giving the number of complaints for the ith doctor, the following explanatory variables were recorded:

   - `residency`, (R), a binary variable taking values N or Y corresponding to whether the doctor had completed residency training;
   - `pay`, (P), giving the dollars per hour earned by the doctor;
   - `hours`, (H), giving the total number of hours worked by the doctor that year.

   The number n_i of visits for the ith doctor over the study period was also recorded.

   One of the models considered was a Poisson regression for the number Y_i of complaints, with residency as the only explanatory variable. Abbreviated R output from this model is shown below.

```
glm(formula = complaints ~ residency + offset(log(visits)), family = poisson)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.4525     0.1026 -62.891   <2e-16 ***
residencyY   -0.3041     0.1725  -1.763   0.0779 .

    Null deviance: 63.435  on 43  degrees of freedom
Residual deviance: 60.245  on 42  degrees of freedom
AIC: 187.03

Number of Fisher Scoring iterations: 5
```

   (a) Explain the role of the offset in the model. **[2 MARKS]**

   (b) Interpret the coefficient of residency in terms of an appropriate rate ratio. Is the effect of residency significant? Explain why. **[4 MARKS]**

   (c) In the above output, what does "Fisher Scoring iterations" refer to? **[2 MARKS]**

   (d) A series of models was fitted to the number of complaints, where H refers to `hours`, P to `pay` and R to `residency`, while H*R is the interaction between H and R and so on. The results are summarised below.

CONTINUED OVERLEAF/

Model	Deviance
Null	63.435
H	57.347
H + P	57.131
H + P + R	55.341
H + P + R + H*P	53.789
H + P + R + H*R	50.182
H + P + R + H*P + H*R	44.747
H + P + R + H*P + H*R + P*R	44.405

Which of the above models would you select, based on the information provided? Justify your choice. **[6 MARKS]**

(e) Explain what is meant by overdispersion, giving two possible causes of overdispersion in a Poisson regression model for counts. Also describe two ways in which we can deal with overdispersion. **[6 MARKS]**

CONTINUED OVERLEAF/

4. For a sample of 53 prostate cancer patients, a number of variables were measured before surgery. The patients then had surgery to determine nodal involvement, i.e. whether the cancer had spread to the lymph nodes. For 20 patients, presence of nodal involvement was confirmed after surgery, while for the remaining 33 patients the absence of nodal involvement was confirmed. The following analysis aims to explore whether nodal involvement could be accurately predicted from the available explanatory variables. The response, `resp`, takes the value 1 for nodal involvement and 0 otherwise. The explanatory variables are:

   - `age`: The patient's age group – 0 for age < 60 and 1 for 60 or over.
   - `stage`: Measurement of the size and position of the tumour observed by palpation with the fingers. A serious case is coded as 1 and a less serious case as 0.
   - `grade`: Another indicator of the seriousness of the cancer which is determined by a pathology reading of a biopsy taken by needle before surgery. A value of 1 indicates a more serious case of cancer and 0 indicates a less serious case.
   - `xray`: Another measure of the seriousness of the cancer taken from an X-ray reading. A value of 1 indicates a more serious case of cancer and 0 indicates a less serious case.
   - `acid`: The level of acid phosphatase in the blood serum where 1=high and 0=low.

   The following abbreviated output is from a logistic regression model fitted to the data to predict nodal involvement:

```
glm(formula = resp ~ age + grade + stage + xray + acid, family = binomial)

Coefficients:
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -3.079     0.987   -3.12      0.0018
age          -0.292     0.754   -0.39      0.6988
grade         0.872     0.816    1.07      0.2850
stage         1.373     0.784    1.75      0.0799
xray          1.801     0.810    2.22      0.0263
acid          1.684     0.791    2.13      0.0334

Null deviance: 70.252 on 52 degrees of freedom
Residual deviance: 47.611 on 47 degrees of freedom
AIC: 59.61

Number of Fisher Scoring iterations: 5
```

   (a) Write down the likelihood for the null model, and show that the maximum likelihood estimator for the probability, p, of nodal involvement, equals y/n where n is the total number of patients in the study and y is the number of patients

CONTINUED OVERLEAF/

with nodal involvement. Use this result to show that the maximised log-likelihood under the null model is given by

$$y \log y + (n - y) \log(n - y) - n \log n$$

and hence verify that the null deviance equals 70.25. **[7 MARKS]**

(b) According to this model, what is the lowest predicted probability of nodal involvement for any future patient?

[4 MARKS]

(c) Which additional model(s)/analysis steps would you consider next for these data? **[3 MARKS]**

(d) The table below shows some of the fitted probabilities from one of the models considered for the prostate cancer data. Describe how you would use these probabilities to predict nodal involvement and how you would evaluate the model's predictive performance.

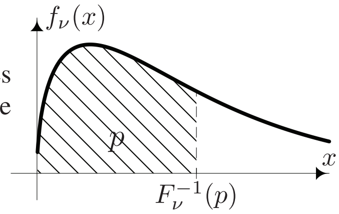
	Resp	Fitted probability $\hat{p}_i$
11	0.00	0.07
12	0.00	0.11
13	0.00	0.13
14	1.00	0.51
15	0.00	0.05
48	1.00	0.22
49	1.00	0.37
50	1.00	0.54
51	1.00	0.88
52	1.00	0.84
53	1.00	0.96

[6 MARKS]

END OF QUESTION PAPER.

5 χ^2 distribution

Inverse $F_\nu^{-1}(p)$ of the cumulative distribution function (quantiles)



The table below contains the quantiles of the χ^2 (chi-squared) distribution with ν degrees of freedom. For $0 < p < 1$ the quantile is the value of x for which $\mathbb{P}\{X \leq x\} = p$, where $X \sim \chi^2(\nu)$. Thus $x = F_\nu^{-1}(p)$.

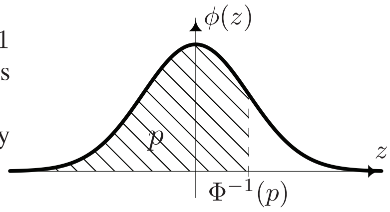
	p											
ν	0.005	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99	0.995	0.999
1	0.0000	0.0002	0.0010	0.0039	0.0158	0.4549	2.7055	3.8415	5.0239	6.6349	7.8794	10.828
2	0.0100	0.0201	0.0506	0.1026	0.2107	1.3863	4.6052	5.9915	7.3778	9.2103	10.597	13.816
3	0.0717	0.1148	0.2158	0.3518	0.5844	2.3660	6.2514	7.8147	9.3484	11.345	12.838	16.266
4	0.2070	0.2971	0.4844	0.7107	1.0636	3.3567	7.7794	9.4877	11.143	13.277	14.860	18.467
5	0.4117	0.5543	0.8312	1.1455	1.6103	4.3515	9.2364	11.070	12.833	15.086	16.750	20.515
6	0.6757	0.8721	1.2373	1.6354	2.2041	5.3481	10.645	12.592	14.449	16.812	18.548	22.458
7	0.9893	1.2390	1.6899	2.1673	2.8331	6.3458	12.017	14.067	16.013	18.475	20.278	24.322
8	1.3444	1.6465	2.1797	2.7326	3.4895	7.3441	13.362	15.507	17.535	20.090	21.955	26.124
9	1.7349	2.0879	2.7004	3.3251	4.1682	8.3428	14.684	16.919	19.023	21.666	23.589	27.877
10	2.1559	2.5582	3.2470	3.9403	4.8652	9.3418	15.987	18.307	20.483	23.209	25.188	29.588
11	2.6032	3.0535	3.8157	4.5748	5.5778	10.341	17.275	19.675	21.920	24.725	26.757	31.264
12	3.0738	3.5706	4.4038	5.2260	6.3038	11.340	18.549	21.026	23.337	26.217	28.300	32.909
13	3.5650	4.1069	5.0088	5.8919	7.0415	12.340	19.812	22.362	24.736	27.688	29.819	34.528
14	4.0747	4.6604	5.6287	6.5706	7.7895	13.339	21.064	23.685	26.119	29.141	31.319	36.123
15	4.6009	5.2293	6.2621	7.2609	8.5468	14.339	22.307	24.996	27.488	30.578	32.801	37.697
16	5.1422	5.8122	6.9077	7.9616	9.3122	15.338	23.542	26.296	28.845	32.000	34.267	39.252
17	5.6972	6.4078	7.5642	8.6718	10.085	16.338	24.769	27.587	30.191	33.409	35.718	40.790
18	6.2648	7.0149	8.2307	9.3905	10.865	17.338	25.989	28.869	31.526	34.805	37.156	42.312
19	6.8440	7.6327	8.9065	10.117	11.651	18.338	27.204	30.144	32.852	36.191	38.582	43.820
20	7.4338	8.2604	9.5908	10.851	12.443	19.337	28.412	31.410	34.170	37.566	39.997	45.315
21	8.0337	8.8972	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932	41.401	46.797
22	8.6427	9.5425	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289	42.796	48.268
23	9.2604	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638	44.181	49.728
24	9.8862	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980	45.559	51.179
25	10.520	11.524	13.120	14.611	16.473	24.337	34.382	37.652	40.646	44.314	46.928	52.620
26	11.160	12.198	13.844	15.379	17.292	25.336	35.563	38.885	41.923	45.642	48.290	54.052
27	11.808	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.195	46.963	49.645	55.476
28	12.461	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278	50.993	56.892
29	13.121	14.256	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588	52.336	58.301
30	13.787	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892	53.672	59.703
31	14.458	15.655	17.539	19.281	21.434	30.336	41.422	44.985	48.232	52.191	55.003	61.098
32	15.134	16.362	18.291	20.072	22.271	31.336	42.585	46.194	49.480	53.486	56.328	62.487
33	15.815	17.074	19.047	20.867	23.110	32.336	43.745	47.400	50.725	54.776	57.648	63.870
34	16.501	17.789	19.806	21.664	23.952	33.336	44.903	48.602	51.966	56.061	58.964	65.247
35	17.192	18.509	20.569	22.465	24.797	34.336	46.059	49.802	53.203	57.342	60.275	66.619
36	17.887	19.233	21.336	23.269	25.643	35.336	47.212	50.998	54.437	58.619	61.581	67.985
37	18.586	19.960	22.106	24.075	26.492	36.336	48.363	52.192	55.668	59.893	62.883	69.346
38	19.289	20.691	22.878	24.884	27.343	37.335	49.513	53.384	56.896	61.162	64.181	70.703
39	19.996	21.426	23.654	25.695	28.196	38.335	50.660	54.572	58.120	62.428	65.476	72.055
40	20.707	22.164	24.433	26.509	29.051	39.335	51.805	55.758	59.342	63.691	66.766	73.402
41	21.421	22.906	25.215	27.326	29.907	40.335	52.949	56.942	60.561	64.950	68.053	74.745
42	22.138	23.650	25.999	28.144	30.765	41.335	54.090	58.124	61.777	66.206	69.336	76.084
43	22.859	24.398	26.785	28.965	31.625	42.335	55.230	59.304	62.990	67.459	70.616	77.419
44	23.584	25.148	27.575	29.787	32.487	43.335	56.369	60.481	64.201	68.710	71.893	78.750
45	24.311	25.901	28.366	30.612	33.350	44.335	57.505	61.656	65.410	69.957	73.166	80.077
46	25.041	26.657	29.160	31.439	34.215	45.335	58.641	62.830	66.617	71.201	74.437	81.400
47	25.775	27.416	29.956	32.268	35.081	46.335	59.774	64.001	67.821	72.443	75.704	82.720
48	26.511	28.177	30.755	33.098	35.949	47.335	60.907	65.171	69.023	73.683	76.969	84.037

→

ν	0.005	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99	0.995	0.999
49	27.249	28.941	31.555	33.930	36.818	48.335	62.038	66.339	70.222	74.919	78.231	85.351
50	27.991	29.707	32.357	34.764	37.689	49.335	63.167	67.505	71.420	76.154	79.490	86.661
51	28.735	30.475	33.162	35.600	38.560	50.335	64.295	68.669	72.616	77.386	80.747	87.968
52	29.481	31.246	33.968	36.437	39.433	51.335	65.422	69.832	73.810	78.616	82.001	89.272
53	30.230	32.018	34.776	37.276	40.308	52.335	66.548	70.993	75.002	79.843	83.253	90.573
54	30.981	32.793	35.586	38.116	41.183	53.335	67.673	72.153	76.192	81.069	84.502	91.872
55	31.735	33.570	36.398	38.958	42.060	54.335	68.796	73.311	77.380	82.292	85.749	93.168
56	32.490	34.350	37.212	39.801	42.937	55.335	69.919	74.468	78.567	83.513	86.994	94.461
57	33.248	35.131	38.027	40.646	43.816	56.335	71.040	75.624	79.752	84.733	88.236	95.751
58	34.008	35.913	38.844	41.492	44.696	57.335	72.160	76.778	80.936	85.950	89.477	97.039
59	34.770	36.698	39.662	42.339	45.577	58.335	73.279	77.931	82.117	87.166	90.715	98.324
60	35.534	37.485	40.482	43.188	46.459	59.335	74.397	79.082	83.298	88.379	91.952	99.607
61	36.301	38.273	41.303	44.038	47.342	60.335	75.514	80.232	84.476	89.591	93.186	100.89
62	37.068	39.063	42.126	44.889	48.226	61.335	76.630	81.381	85.654	90.802	94.419	102.17
63	37.838	39.855	42.950	45.741	49.111	62.335	77.745	82.529	86.830	92.010	95.649	103.44
64	38.610	40.649	43.776	46.595	49.996	63.335	78.860	83.675	88.004	93.217	96.878	104.72
65	39.383	41.444	44.603	47.450	50.883	64.335	79.973	84.821	89.177	94.422	98.105	105.99
66	40.158	42.240	45.431	48.305	51.770	65.335	81.085	85.965	90.349	95.626	99.330	107.26
67	40.935	43.038	46.261	49.162	52.659	66.335	82.197	87.108	91.519	96.828	100.55	108.53
68	41.713	43.838	47.092	50.020	53.548	67.335	83.308	88.250	92.689	98.028	101.78	109.79
69	42.494	44.639	47.924	50.879	54.438	68.334	84.418	89.391	93.856	99.228	103.00	111.06
70	43.275	45.442	48.758	51.739	55.329	69.334	85.527	90.531	95.023	100.43	104.21	112.32
71	44.058	46.246	49.592	52.600	56.221	70.334	86.635	91.670	96.189	101.62	105.43	113.58
72	44.843	47.051	50.428	53.462	57.113	71.334	87.743	92.808	97.353	102.82	106.65	114.84
73	45.629	47.858	51.265	54.325	58.006	72.334	88.850	93.945	98.516	104.01	107.86	116.09
74	46.417	48.666	52.103	55.189	58.900	73.334	89.956	95.081	99.678	105.20	109.07	117.35
75	47.206	49.475	52.942	56.054	59.795	74.334	91.061	96.217	100.84	106.39	110.29	118.60
76	47.997	50.286	53.782	56.920	60.690	75.334	92.166	97.351	102.00	107.58	111.50	119.85
77	48.788	51.097	54.623	57.786	61.586	76.334	93.270	98.484	103.16	108.77	112.70	121.10
78	49.582	51.910	55.466	58.654	62.483	77.334	94.374	99.617	104.32	109.96	113.91	122.35
79	50.376	52.725	56.309	59.522	63.380	78.334	95.476	100.75	105.47	111.14	115.12	123.59
80	51.172	53.540	57.153	60.391	64.278	79.334	96.578	101.88	106.63	112.33	116.32	124.84
81	51.969	54.357	57.998	61.261	65.176	80.334	97.680	103.01	107.78	113.51	117.52	126.08
82	52.767	55.174	58.845	62.132	66.076	81.334	98.780	104.14	108.94	114.69	118.73	127.32
83	53.567	55.993	59.692	63.004	66.976	82.334	99.880	105.27	110.09	115.88	119.93	128.56
84	54.368	56.813	60.540	63.876	67.876	83.334	100.98	106.39	111.24	117.06	121.13	129.80
85	55.170	57.634	61.389	64.749	68.777	84.334	102.08	107.52	112.39	118.24	122.32	131.04
86	55.973	58.456	62.239	65.623	69.679	85.334	103.18	108.65	113.54	119.41	123.52	132.28
87	56.777	59.279	63.089	66.498	70.581	86.334	104.28	109.77	114.69	120.59	124.72	133.51
88	57.582	60.103	63.941	67.373	71.484	87.334	105.37	110.90	115.84	121.77	125.91	134.75
89	58.389	60.928	64.793	68.249	72.387	88.334	106.47	112.02	116.99	122.94	127.11	135.98
90	59.196	61.754	65.647	69.126	73.291	89.334	107.57	113.15	118.14	124.12	128.30	137.21
91	60.005	62.581	66.501	70.003	74.196	90.334	108.66	114.27	119.28	125.29	129.49	138.44
92	60.815	63.409	67.356	70.882	75.100	91.334	109.76	115.39	120.43	126.46	130.68	139.67
93	61.625	64.238	68.211	71.760	76.006	92.334	110.85	116.51	121.57	127.63	131.87	140.89
94	62.437	65.068	69.068	72.640	76.912	93.334	111.94	117.63	122.72	128.80	133.06	142.12
95	63.250	65.898	69.925	73.520	77.818	94.334	113.04	118.75	123.86	129.97	134.25	143.34
96	64.063	66.730	70.783	74.401	78.725	95.334	114.13	119.87	125.00	131.14	135.43	144.57
97	64.878	67.562	71.642	75.282	79.633	96.334	115.22	120.99	126.14	132.31	136.62	145.79
98	65.694	68.396	72.501	76.164	80.541	97.334	116.32	122.11	127.28	133.48	137.80	147.01
99	66.510	69.230	73.361	77.046	81.449	98.334	117.41	123.23	128.42	134.64	138.99	148.23
100	67.328	70.065	74.222	77.929	82.358	99.334	118.50	124.34	129.56	135.81	140.17	149.45
110	75.550	78.458	82.867	86.792	91.471	109.33	129.39	135.48	140.92	147.41	151.95	161.58
120	83.852	86.923	91.573	95.705	100.62	119.33	140.23	146.57	152.21	158.95	163.65	173.62
150	109.14	112.67	117.98	122.69	128.28	149.33	172.58	179.58	185.80	193.21	198.36	209.26
200	152.24	156.43	162.73	168.28	174.84	199.33	226.02	233.99	241.06	249.45	255.26	267.54
500	422.30	429.39	439.94	449.15	459.93	499.33	540.93	553.13	563.85	576.49	585.21	603.45

Inverse $\Phi^{-1}(p)$ of the cumulative distribution function (quantiles)

The table below contains the quantiles of the standard normal distribution. For $0 < p < 1$ the quantile is the value of z for which $\mathbb{P}\{Z \leq z\} = p$, where $Z \sim \mathsf{N}(0,1)$. Thus $z = \Phi^{-1}(p)$.
The table only contains the quantiles for $p \geq \frac{1}{2}$. For $p < \frac{1}{2}$ quantiles can be obtained by exploiting the symmetry of the normal distribution: $\Phi^{-1}(p) = -\Phi^{-1}(1-p)$.



	p									
p	0.□□0	0.□□1	0.□□2	0.□□3	0.□□4	0.□□5	0.□□6	0.□□7	0.□□8	0.□□9
0.50□	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51□	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52□	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53□	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54□	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231
0.55□	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56□	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57□	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	0.1968	0.1993
0.58□	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59□	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60□	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61□	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62□	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63□	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64□	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65□	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66□	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67□	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68□	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69□	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70□	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71□	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72□	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73□	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74□	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713
0.75□	0.6745	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76□	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77□	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78□	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79□	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80□	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8705	0.8742
0.81□	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82□	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83□	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84□	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85□	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86□	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87□	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88□	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89□	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90□	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91□	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92□	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93□	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94□	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95□	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96□	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97□	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98□	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99□	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

Selected quantiles $\Phi^{-1}(p)$ in high precision

p	$\Phi^{-1}(p)$
0.9	1.2815515655
0.95	1.6448536270
0.975	1.9599639845
0.99	2.3263478740
0.995	2.5758293035
0.9975	2.8070337683

p	$\Phi^{-1}(p)$
0.999	3.0902323062
0.9995	3.2905267315
0.99975	3.4807564043
0.9999	3.7190164855
0.99995	3.8905918864
0.999975	4.0556269811

p	$\Phi^{-1}(p)$
0.99999	4.2648907939
0.999995	4.4171734135
0.9999975	4.5647877303
0.999999	4.7534243088
0.9999995	4.8916384757
0.99999975	5.0263128360

Discrete univariate distributions

Distribution	Probability mass function (p.m.f.) $p(x)$	Range	Parameters	$\mathbb{E}(X)$	$\text{Var}(X)$	Moment-generating function $M(t)$	Comments
Binomial[1,2] $\text{Bi}(n,\theta)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$	$x \in \{0,1,\ldots,n\}$	$n \in \mathbb{N}$, $0<\theta<1$	$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta\exp(t))^n$	No. of successes in n trials; θ – probability of success
Geometric $\text{Geo}(\theta)$	$\theta^{x-1}(1-\theta)$	$x \in \mathbb{N}$	$0<\theta<1$	$\frac{1}{1-\theta}$	$\frac{\theta}{(1-\theta)^2}$	$\frac{(1-\theta)\exp(t)}{1-\theta\exp(t)}$	No. of trials until (and including) first failure; θ – probability of success
Hypergeometric $\text{HyGe}(n,N,M)$	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$x\in\{\max\{0,n-(N-M)\},\ldots,\min\{n,M\}\}$	$N,n \in \mathbb{N}$, $M \in \{0,\ldots,N\}$	$n\theta$	$n\theta(1-\theta)\cdot\frac{N-n}{N-1}$ (with $\theta=\frac{M}{N}$)	—[3]	No. of type I objects in a sample of size n, drawn *without* replacement from a population of size N, containing M type I objects.
Negative Binomial $\text{NeBi}(k,\theta)$	$\binom{x-1}{k-1}\theta^{x-k}(1-\theta)^k$	$x \in \{k,k+1,\ldots\}$	$k \in \mathbb{N}$, $0<\theta<1$	$\frac{k}{1-\theta}$	$\frac{k\theta}{(1-\theta)^2}$	$\left(\frac{(1-\theta)\exp(t)}{1-\theta\exp(t)}\right)^k$	No. of trials until (and including) k^{th} failure; θ – probability of success; $\text{NeBi}(1,\theta)\equiv\text{Geo}(\theta)$
Poisson $\text{Poi}(\lambda)$	$\exp(-\lambda)\frac{\lambda^x}{x!}$	$x \in \mathbb{N}_0$	$\lambda>0$	λ	λ	$\exp(\lambda(\exp(t)-1))$	

[1] $\text{Bi}(n,\theta)$ can be approximated by $\text{Poi}(n\theta)$, if n large, θ small and $n\theta$ moderate.

[2] $\text{Bi}(n,\theta)$ can be approximated by $\text{N}(n\theta, n\theta(1-\theta))$, if n large and θ not too close to 0 or 1.

[3] No simple closed form expression exists.

Continuous univariate distributions

Distribution	Probability density function (p.d.f.) $f(x)$	Range	Parameters	$\mathbb{E}(X)$	$\text{Var}(X)$	Moment-generating function $M(t)$	Comments
Beta $\text{Be}(\alpha_1,\alpha_2)$	$\frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{\text{B}(\alpha_1,\alpha_2)}$	$0\le x\le 1$	$\alpha_1>0$, $\alpha_2>0$	$\frac{\alpha_1}{\alpha_1+\alpha_2}$	$\frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}$	—[3]	$X_1\sim\text{Ga}(\alpha_1,\theta)$, $X_2\sim\text{Ga}(\alpha_2,\theta)$ independent $\Rightarrow \frac{X_1}{X_1+X_2}\sim\text{Be}(\alpha_1,\alpha_2)$
Cauchy $\text{Ca}(\eta,\gamma)$	$\frac{1}{\pi\gamma\left(1+\frac{(x-\eta)^2}{\gamma^2}\right)}$	$x\in\mathbb{R}$	$\eta\in\mathbb{R}$, $\gamma>0$	—[4]	—[4]	—[4]	$\text{Ca}(0,1)\equiv\text{t}(1)$
Chi-Squared $\chi^2(\nu)$	$\frac{x^{\frac{\nu}{2}-1}\exp\left(-\frac{x}{2}\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$	$x>0$	$\nu\in\mathbb{N}$	ν	2ν	$\frac{1}{(1-2t)^{\frac{\nu}{2}}}$	$X_i\sim\text{N}(0,1)$ independent $\Rightarrow \sum_{i=1}^{\nu}X_i^2\sim\chi^2(\nu)$
Exponential $\text{Expo}(\theta)$	$\theta\exp(-\theta x)$	$x>0$	$\theta>0$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$	$\frac{1}{1-\frac{t}{\theta}}$	
F $\text{F}(\nu_1,\nu_2)$	$\frac{\nu_1^{\frac{\nu_1}{2}}\nu_2^{\frac{\nu_2}{2}}}{\text{B}\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}\frac{x^{\frac{\nu_1}{2}-1}}{(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}}$	$x>0$	$\nu_1,\nu_2\in\mathbb{N}$	$\frac{\nu_2}{\nu_2-2}$ (for $\nu_2>2$)	$\frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ (for $\nu_2>4$)	—[4]	$X_1\sim\chi^2(\nu_1)$, $X_2\sim\chi^2(\nu_2)$ independent $\Rightarrow \frac{X_1/\nu_1}{X_2/\nu_2}\sim\text{F}(\nu_1,\nu_2)$

Distribution	Probability density function (p.d.f.) $f(x)$	Range	Parameters	$\mathbb{E}(X)$	$\mathrm{Var}(X)$	Moment-generating function $M(t)$	Comments
Gamma $\mathsf{Ga}(\alpha, \theta)$	$\dfrac{\theta^\alpha x^{\alpha-1}\exp(-\theta x)}{\Gamma(\alpha)}$	$x > 0$	$\theta > 0$ $\alpha > 0$	$\dfrac{\alpha}{\theta}$	$\dfrac{\alpha}{\theta^2}$	$\dfrac{1}{\left(1-\frac{t}{\theta}\right)^\alpha}$	$\mathsf{Ga}(1,\theta) \equiv \mathsf{Expo}(\theta)$ $\mathsf{Ga}\left(\frac{\nu}{2},\frac{1}{2}\right) \equiv \chi^2(\nu)$
Normal $\mathsf{N}(\mu, \sigma^2)$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$	$\mu \in \mathbb{R}$ $\sigma^2 > 0$	μ	σ^2	$\exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$	$\mathsf{N}(0,1)$ – standard normal $\frac{X-\mu}{\sigma} \sim \mathsf{N}(0,1)$
Pareto $\mathsf{Pa}(k, \theta)$	$\dfrac{\theta k^\theta}{x^{\theta+1}}$	$x > k$	$k > 0$ $\theta > 0$	$\dfrac{\theta k}{\theta - 1}$ (for $\theta > 1$)	$\dfrac{\theta k^2}{(\theta-1)^2(\theta-2)}$ (for $\theta > 2$)	—[3]	
Student's t $\mathsf{t}(\nu)$	$\dfrac{1}{\sqrt{\nu}\mathrm{B}\left(\frac{\nu}{2},\frac{1}{2}\right)\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}}$	$x \in \mathbb{R}$	$\nu \in \mathbb{N}$	0 (for $\nu > 1$)	$\dfrac{\nu}{\nu-2}$ (for $\nu > 2$)	—[4]	$X_1 \sim \mathsf{N}(0,1)$ $X_2 \sim \chi^2(\nu)$ independent $\Rightarrow \frac{X_1}{\sqrt{\frac{X_2}{\nu}}} \sim \mathsf{t}(\nu)$
Uniform $\mathsf{U}(a, b)$ c.d.f. for $a < x < b$:	$\dfrac{1}{b-a}$	$a \leq x \leq b$	$a, b \in \mathbb{R}$ $a < b$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{\exp(bt)-\exp(at)}{t(b-a)}$	$\mathsf{U}(0,1) \equiv \mathsf{Be}(1,1)$
Weibull $\mathsf{We}(\alpha, \theta)$	$\alpha\theta x^{\alpha-1}\exp(-\theta x^\alpha)$	$x > 0$	$\alpha > 0$ $\theta > 0$	$\dfrac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\theta^{\frac{1}{\alpha}}}$	$\dfrac{\Gamma\left(1+\frac{2}{\alpha}\right)}{\theta^{\frac{2}{\alpha}}} - (\mathbb{E}(X))^2$	—[3]	$\mathsf{We}(1,\theta) \equiv \mathsf{Expo}(\theta)$

[3] No simple closed form expression exists.

[4] Does not exist.

Multivariate distributions

Distribution	Probability mass / density function $p(\mathbf{x}) = p(x_1, \dots, x_k)$ or $f(\mathbf{x}) = f(x_1, \dots, x_k)$	Range	Parameters	$\mathbb{E}(X_j)$	$\mathrm{Var}(X_j)$	$\mathrm{Cov}(X_i, X_j)$	Moment-generating function $M(\mathbf{t}) = M(t_1, \dots, t_k)$
Multinomial $\mathsf{Mu}(n, \theta_1, \dots \theta_k)$	$\dfrac{n!}{x_1! \cdots x_k!}\theta_1^{x_1} \cdots \theta_k^{x_k}$	$x \in \mathbb{N}_0$ $\sum_{j=1}^k x_j = n$	$n \in \mathbb{N}$ $0 < \theta_j < 1$ $\sum_{j=1}^k \theta_j = 1$	$n\theta_j$	$n\theta_j(1-\theta_j)$	$-n\theta_i\theta_j$	$\left(\displaystyle\sum_{j=1}^k \theta_j \exp(t_j)\right)^n$
Normal $\mathsf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\dfrac{1}{(2\pi)^{\frac{k}{2}}\lvert\boldsymbol{\Sigma}\rvert^{1/2}}\exp\left(-\dfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$	$x \in \mathbb{R}^k$	$\boldsymbol{\mu} \in \mathbb{R}^k$ $\boldsymbol{\Sigma}$ symm., pos. def.	μ_j	Σ_{jj}	Σ_{ij}	$\exp\left(\boldsymbol{\mu}^\top\mathbf{t} + \dfrac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}\right)$
Special case: bivariate normal distribution with $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$:							
$\dfrac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}\exp\left(-\dfrac{\sigma_2^2(x_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right)$			$\mu_1, \mu_2 \in \mathbb{R}$ $\sigma_1^2, \sigma_2^2 > 0$ $-1 < \rho < 1$	μ_j	σ_j^2	$\rho\sigma_1\sigma_2$	