CS470, Fall 2017
Homework 4. Basic Probability

TA: Geon-Hyeong Kim (ghkim@ai.kaist.ac.kr)

Nov. 22 - Dec. 1, 2017

1 Backgrounds

The Bernoullli distribution $\mathrm{Bern}(p)$ is the discrete probability distribution of a random variable that takes the value 1 with probability p and the value 0 with probability $1-p$. Thus, we can write the probaility mass function of it as

$$\mathrm{Bern}(x;p) = p^x(1-p)^{1-x}$$

with support $\{0,1\}$.

The binomial distribution $\mathrm{B}(n,p)$ is the discrete probability distribution of the number of successes in a sequence of n independent trials with success probability p for each trial. Thus, Bernoulli distribution is a special case of binomial distribution with $n=1$. The probability mass function of binomial distribution is written by

$$\mathrm{B}(x;n,p) = \binom{n}{x} p^x(1-p)^{n-x}$$

with support $\{0,1,\ldots,n\}$.

The Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$ is a widely used continuous probability distribution which has mean μ and variance σ^2. The probability density function of it can be written in the form

$$\mathcal{N}(x;\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

with support $\mathbb{R}$.

For a D-dimensional vector $\mathbf{x}$, the probability densify function of multivariate Gaussian distribution takes the form

$$\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

where $\boldsymbol{\mu}$ is a D-dimensional mean vector, $\boldsymbol{\Sigma}$ is a $D\times D$ covariance matrix, $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$.

2 Problems

For each problem without any additional mention, you **must write a solution** as well as an answer. The following problems are based on Bishop [2006].

Problem 1. [20 points]

(a) **[10 points]** Derive the mean and variance of Bernoulli distribution $\mathrm{Bern}(x;p)$.

(b) **[10 points]** Derive the mean and variance of binomial distribution $\mathrm{B}(x;n,p)$.

Problem 2. [50 points]

(a) **[15 points]** For D-dimensional vector $\mathbf{x}$, derive the mean and variance of multivariate Gaussian distribution $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$.

(b) **[10 points]** Suppose that the matrix $\mathbf{M}$ can be partitioned by

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{D}$ is invertible. Then, we can derive the inverse of $\mathbf{M}$ as partitioned matrix

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{A}' & \mathbf{B}' \\ \mathbf{C}' & \mathbf{D}' \end{pmatrix},$$

where each block matrix $\mathbf{M}^{-1}$ has same size of corresponding block matrix of $\mathbf{M}$. Derive the block matrices of $\mathbf{M}^{-1}$ using the block matrices of $\mathbf{M}$.

(c) **[15 points]** Let M be the integer which satisfies $0 < M < D$. Then, we can take M-dimensional vector $\mathbf{x}_a$ to form the first M components of $\mathbf{x}$ with $\mathbf{x}_b$ comprising the remaining $D - M$ compontnents, so that

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}.$$

We also define corresponding partitions of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}.$$

Using these notations and the result of (b), derive the mean and variance of the conditional distribution $p(\mathbf{x}_a|\mathbf{x}_b)$.

(d) **[10 points]** Derive the mean and variance of the marginal distribution $p(\mathbf{x}_a)$.

Problem 3. [30 points] Using the Fig. 1 and the Gaussian distribution $\mathcal{N}(x; \mu, \sigma^2)$, solve the following problems.

(a) **[15 points]** Let $\mu = 5$ and $\sigma = 3$. Compute the probability that $x \in [2.3, 8.3]$.

(b) **[15 points]** Let $\mu = 10$ and $\sigma = 5$. Compute the $y \in \mathbb{R}_+$ which satisfies the following:

$$\Pr\big(x \in [10 - y, 10 + y]\big) = 0.95$$

References

Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

Wikipedia. *Standard normal table*, 2017 (accessed Nov 20, 2017). `https://en.wikipedia.org/wiki/Standard_normal_table`.

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670

Figure 1: Cumulative of standard Gaussian distribution from mean (0 to Z). Wikipedia [2017 (accessed Nov 20, 2017]