18.650. Fundamentals of Statistics
Fall 2023. Recitation sheet 1.2

# 1 Multivariate random variables and limit theorems

Problem 1 Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left( \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

Compute the following quantities

1. $\mathbb{V}[X], \mathbb{V}[Y]$
2. $\mathbb{E}[(X-Y)^2]$
3. $\mathbb{V}[X+2Y]$
4. $\mathbb{E}[X^2Y]$. (Hint: find a number $a$ such that $Y - aX$ is uncorrelated with $X$)

Problem 2 As usual, $\mathbb{E}X := (\mathbb{E}X_1, \ldots, \mathbb{E}X_d)^\top$ for a random vector $X \in \mathbb{R}^d$.

1. Let $X \in \mathbb{R}^d$ be a random vector and let $L : \mathbb{R}^d \to \mathbb{R}^k$ be a linear map. Prove that $\mathbb{E}[L(X)] = L(\mathbb{E}[X])$.

2. Let $Y$ be a random $d \times d$ matrix and $A$ be a deterministic $d \times d$ matrix. Prove that $\mathbb{E}Tr(AY) = Tr(A\mathbb{E}[Y])$. (Recall that the trace $Tr$ of a square matrix is the sum of its diagonal entries.)

3. Let $X \in \mathbb{R}^d$ be a random vector such that $\mathbb{E}[X] = (0, \ldots, 0)^\top$ and $\mathbb{V}[X] = I_d$. Let $A$ be a deterministic $d \times d$ matrix. Compute $\mathbb{E}[X^\top AX]$.

Problem 3 (AoS Exercise 5.15) Let

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix}, \ldots, \begin{pmatrix} X_{1n} \\ X_{2n} \end{pmatrix}$$

be i.i.d. random vectors with mean $\mu = (\mu_1, \mu_2) = (1, -1)$ and variance $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$.
Let

$$\bar{X}_{n,1} = \frac{1}{n}\sum_{i=1}^{n} X_{1i}, \quad \bar{X}_{n,2} = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$$

and define $Y_n = \bar{X}_{n,1}/\bar{X}_{n,2}$. Find the limiting distribution of $\sqrt{n}(Y_n + 1)$.

## 2 Parameter estimation and MSE

Problem 4 Let $X_1, \dots, X_n$ be i.i.d. $\mathsf{Unif}(0, \theta)$ variables for some unknown $\theta > 0$.

1. Write down a valid statistical model for the resulting data.
2. Compute the cdf, pdf, expectation, and variance of $\max_i X_i$.
3. Suppose we use $\hat{\theta} = \max_i X_i$ as an estimator for $\theta$. Compute the MSE of this estimator.
4. Find $a > 0$ such that $a \max_i X_i$ is an unbiased estimator of $\theta$ and compute its MSE. How does it compare to the MSE of $\hat{\theta} = \max_i X_i$?

Problem 5 Let $X_1, \dots, X_n \sim N(\theta, 1)$. Define

$$Y_i = \begin{cases} 1 & \text{if } X_i > 0, \\ 0 & \text{if } X_i \leq 0. \end{cases}$$

We're interested in estimating $\psi = \mathsf{P}(Y_1 = 1)$.

1. Find $f$ such that $\psi = f(\theta)$.
2. This motivates using the "plug-in" estimator $\hat{\psi} := f(\bar{X}_n)$ of $\psi$. Find $\mathrm{se}(\hat{\psi})$.
3. Construct an approximate 95% confidence interval for $\psi$.
4. Let $\bar{\psi} = \bar{Y}_n$ be another estimator of $\psi$. Find $\mathrm{se}(\bar{\psi})$
5. Suppose $\theta = 0$. Which is smaller: $\mathrm{se}(\bar{\psi})$ or $\mathrm{se}(\hat{\psi})$?

## 3 Maximum Likelihood Estimators

Problem 6 In each of the following cases write the likelihood function and compute the maximum likelihood estimator of the parameter based on the $X_i$:

1. $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$.
2. $X_1, \dots, X_n$ are i.i.d. $\mathrm{Pois}(\lambda)$ random variables.

3. $X_1, \ldots, X_n$ are i.i.d. $\mathsf{Exp}(\lambda)$ random variables.
4. $X_1, \ldots, X_n$ are i.i.d. Unif$([0, \theta])$ random variables.
5. $X_1, \ldots, X_n$ are i.i.d. Bernoulli$(p)$

**Definition 1** *(Fisher Information for Random Variables) Let $X$ be a random variable with pdf $f_\theta(x)$ where $\theta$ is some parameter. The Fisher Information $I(\theta)$ of $\theta$ is defined by:*

$$I(\theta) = V_\theta \left( \frac{\partial \log f_\theta(x)}{\partial \theta} \right)$$

*Under certain regularity conditions, one can show that this is equivalent to:*

$$I(\theta) = -\mathsf{E}_\theta \left[ \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} \right] = -\int \frac{\partial^2 \log f_\theta(x)}{\partial \theta^2} f_\theta(x) dx$$

**Theorem 1** *(Asymptotic Normality of MLE) Let $\hat{\theta}_n$ be the maximum likelihood estimator of $\theta$. Then under appropriate regularity conditions:*

$$\sqrt{n}(\hat{\theta}_n - \theta) \to^d N(0, I(\theta)^{-1})$$

*Amongst these regularity conditions is the condition that the support of $f_\theta(x)$ does not depend on $\theta$.*

Problem 7 Consider the maximum likelihood estimators evaluated in parts 1-4 of problem 6.

1. For each part determine whether or not the corresponding Fisher information is well-defined.
2. For each part where the Fisher information is well-defined determine it and the asymptotic variance of the estimator.

Problem 8 Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Let $\tau$ be such that $\mathsf{P}(X < \tau) = 0.95$.

1. Write $\tau$ in terms of $\mu$ and $\sigma$.
2. Write down an estimator $\hat{\tau}$ of $\tau$ in terms of the MLEs $\hat{\mu}, \hat{\sigma}$ for $\mu$ and $\sigma$, respectively.
3. Find an expression for an approximate $1 - \alpha$ confidence interval for $\tau$. Hint: use the asymptotic variance of $\hat{\mu}, \hat{\sigma}$ computed in part 2 of problem 7. Then apply the delta method to get the asymptotic variance of $\tau$.

# 4 Method of Moments and EM

**Problem 9** The gamma distribution is given by

$$f_{k,\theta}(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta},$$

where $\Gamma(k)$ is the gamma function. The parameter $k$ is known as a shape parameter, and $\theta$ is called the scale parameter. The first and second moments are

$$\begin{aligned} \alpha_1 &= \mathsf{E}_{k,\theta}[X] = k\theta, \\ \alpha_2 &= \mathsf{E}_{k,\theta}[X^2] = (k+k^2)\theta^2 \end{aligned} \tag{1}$$

Given $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_{k,\theta}$ find the method of moments estimators $\hat{k}, \hat{\theta}$

**Problem 10** Let $X_1, \ldots, X_n$ be i.i.d. with pdf

$$f_\mu(x) = \frac{1}{2}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}\right] + \frac{1}{2}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x+\mu)^2}{2}}\right],$$

where $\mu \geq 0$. This is an equally-weighted mixture of the two Gaussian distributions $\mathcal{N}(\mu, 1)$ and $\mathcal{N}(-\mu, 1)$.

1. What is the log likelihood $\ell_n(\mu)$?
2. Let $Y_i \in \{-1, 1\}$ be the hidden variable telling us the membership of $X_i$ to either $\mathcal{N}(-\mu, 1)$ or $\mathcal{N}(\mu, 1)$. What is the log likelihood $\ell_n(\mu)$ in the case that we observe both the $X_i$'s and the $Y_i$'s? What is the MLE for $\mu$ in this case?
3. Let $\mu_k$ be our current guess for $\mu$. Compute $\hat{Y}_i = \mathsf{E}_{\mu_k}[Y_i \mid X_i]$. This is the E-step of the EM algorithm.) Use this to construct an approximation $\hat{\ell}_n(\mu)$ to the log likelihood in the observed-label case.
4. Find $\mu_{k+1} = \operatorname{argmax}_\mu \hat{\ell}_n(\mu)$ to get an update rule $\mu_k \mapsto \mu_{k+1}$. (This is the M step of the EM algorithm.) Check that this rule makes sense intuitively, and compare it to the estimator $\hat{\mu}$ from part 1.
5. What happens if you initialize the algorithm at $\mu_0 = 0$?
6. Compute $\mathsf{E}[X_1^2]$ in terms of $\mu$. Based on this, construct the method of moments estimator $\hat{\mu}$ of $\mu$.
7. Suppose $\mu = 1$. What is the asymptotic variance of $\hat{\mu}$?

| | Second decimal place of $Z$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $Z$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |

The table lists $P(Z \leq z)$ where $Z \sim N(0,1)$ for positive values of $z$.