MATH 541a Homework 3

Qilin Ye

February 18, 2022

Problem 1

Let $n \geqslant 2$. Let $S^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$. Let v be a random vector uniformly distributed in S^{n-1}. Prove that for any $t > 0$ and any $x \in S^{n-1}$ fixed,

$$\mathbb{P}(v \in S^{n-1} : |\langle v, x\rangle| > t/\sqrt{n}) \leqslant \frac{10}{t}.$$

Proof. (The first half is identical to homework 1 problem 11.) We first provide an upper bound for $\mathbb{E}|\langle x, v\rangle|$ and then use Markov's inequality to conclude the proof.

Since the uniform distribution on S^{n-1} is invariant under rotations about the origin, and since inner product is also preserved under rotations, i.e., $\langle a, b\rangle = \langle Ra, Rb\rangle$, we have, for any rotation $R : S^{n-1} \to S^{n-1}$,

$$\mathbb{E}|\langle x, v\rangle| = \mathbb{E}|\langle x, Rv\rangle| = \mathbb{E}|\big\langle R^{-1}x, R^{-1}Rv\big\rangle| = \mathbb{E}|\big\langle R^{-1}x, v\big\rangle|.$$

Therefore we can WLOG assume $x = (1, 0, \ldots, 0)$. Then

$$\mathbb{E}|\langle x, v\rangle| = \mathbb{E}|\langle u, v\rangle| = \frac{1}{\text{Area}(S^{n-1})}\int_{S^{n-1}} |v_1|\, \mathrm{d}V. \tag{Q1.1}$$

Note that under spherical coordinates with parameters $r, \varphi_1, \varphi_2, \ldots, \varphi_{n-1}$, the first component v_1 can be expressed as $r\cos\varphi_1$, and the Jacobian is

$$r^{n-1}\prod_{i=1}^{n-2}\sin^{n-1-i}(\varphi_i) = r^{n-1}\sin^{n-2}(\varphi - 1)\sin^{n-3}(\varphi_2)\cdots\sin(\varphi_{n-2}).$$

In this case $r \equiv 1$ on S^{n-1} so we get two simpler $(n-1)$-fold integrals:

$$\int_{S^{n-1}} |v_1|\, \mathrm{d}V = \int_{\varphi_{n-1}=0}^{2\pi}\int_{\varphi_{n-2}=0}^{\pi}\cdots\int_{\varphi_1=0}^{\pi}|\cos\varphi_1|\prod_{i=1}^{n-2}\sin^{n-1-i}(\varphi_i)\, \mathrm{d}\varphi_1\cdots\mathrm{d}\varphi_{n-2}\,\mathrm{d}\varphi_{n-1} \tag{Q1.2}$$

and

$$\text{Area}(S^{n-1}) = \int_{S^{n-1}} 1\, \mathrm{d}V = \int_{\varphi_{n-1}=0}^{2\pi}\int_{\varphi_{n-2}=0}^{\pi}\cdots\int_{\varphi_1=0}^{\pi}\prod_{i=1}^{n-2}\sin^{n-1-i}(\varphi_i)\, \mathrm{d}\varphi_1\cdots\mathrm{d}\varphi_{n-2}\,\mathrm{d}\varphi_{n-1}. \tag{Q1.3}$$

Division gives $\mathbb{E}|\langle x, v\rangle| = (\text{Q1.2})/(\text{Q1.3}) = \int_0^\pi |\cos\varphi|\sin^{n-2}\varphi\, \mathrm{d}\varphi \Big/ \int_0^\pi \sin^{n-2}\varphi\, \mathrm{d}\varphi$. Since both integrals satisfy $\int_0^\pi = 2\int_0^{\pi/2}$, the ratio further equals $\int_0^{\pi/2}\cos\varphi\sin^{n-2}\varphi\, \mathrm{d}\varphi \Big/ \int_0^{\pi/2}\sin^{n-2}\varphi\, \mathrm{d}\varphi$.

The numerator is $1/(n-1) < 1/n$ by a simple u-substitution with $u := \sin\varphi$. For $n \geqslant 2$, we bound the denominator from below by noticing $\cos^{n-2} x \geqslant \cos^n x \geqslant e^{-nx^2}$ on $[0,1]$, which implies

$$\begin{aligned}\int_0^{\pi/2} \sin^{n-2}\varphi \,\mathrm{d}\varphi &= \int_0^{\pi/2} \cos^{n-2}\varphi \,\mathrm{d}\varphi \geqslant \int_0^1 \cos^{n-2}\varphi \,\mathrm{d}\varphi \\ &\geqslant \int_0^1 e^{-nx^2}\,\mathrm{d}x = \frac{1}{\sqrt{n}}\int_0^{\sqrt{n}} e^{-u^2}\,\mathrm{d}u \geqslant \frac{1}{\sqrt{n}}\int_0^1 e^{-u^2}\,\mathrm{d}u > \frac{1}{10\sqrt{n}}.\end{aligned}$$

Combining both bounds, we obtain

$$\mathbb{E}|x,v| < \frac{1/n}{1/(10\sqrt{n})} = \frac{10}{\sqrt{n}}.$$

Then the problem is simply an application of Markov's inequality:

$$\mathbb{P}(v : |\langle x,v\rangle| > t/\sqrt{n}) \leqslant \frac{\mathbb{E}|\langle v,x\rangle|}{t/\sqrt{n}} < \frac{10}{t}. \qquad \square$$

Problem 2

Let X be uniformly distributed on $[0,1]$. Show that the location family of X is not an exponential family, i.e., the corresponding densities $\{f(x+\mu)\}_{\mu\in\mathbb{R}}$ cannot be written in the form

$$h(x)\exp((w(\mu)t(x) - a(w(\mu))$$

where $h : \mathbb{R} \to [0,\infty), w : \mathbb{R} \to \mathbb{R}, t : \mathbb{R} \to \mathbb{R}, x \in \mathbb{R}$, and $a(w(\mu))$ the appropriate scaling factor.

Proof. Note that for any give μ, the PDF is zero outside $[\mu, \mu+1]$. Since $\exp(\cdot)$ is nonzero, this means $h \equiv 0$ on $\mathbb{R}\backslash[\mu,\mu+1]$. Letting μ vary, we see h needs to be zero everywhere. Then the PDF is zero, which is absurd. $\square$

Problem 3

Suppose we have a k-parameter exponential family in canonical form so that

$$f_w(x) := h(x)\exp\Big(\sum_{i=1}^k w_i t_i(x) - a(w)\Big) \qquad \text{for all } w \in \mathbb{R}^k, x \in \mathbb{R}^n,$$

$$a(w) := \log \int_{\mathbb{R}^n} h(x)\exp\Big(\sum_{i=1}^k w_i t_i(x)\Big)\,\mathrm{d}\mu(x),$$

and

$$W := \{w \in \mathbb{R}^k : a(w) < \infty\}.$$

Show that $a(w)$ is convex and conclude that W is a convex set.
Hint: use Hölder's inequality $\|fg\|_1 \leqslant \|f\|_p\|g\|_q$ *where* $1/p + 1/q = 1$.

Proof. We begin by picking arbitrary $u, v \in \mathbb{R}^k$. Also, pick $\lambda \in (0,1)$ and consider the conjugate pair $1/\lambda$ and

$1/(1-\lambda)$. Since $h, \exp \geqslant 0$, we may drop the absolute values inside the integrals. From definition, we have

$$\begin{aligned}
a(\lambda u + (1-\lambda)v) &= \log \int_{\mathbb{R}^n} h(x) \exp\Big(\sum_{i=1}^k (\lambda u_i + (1-\lambda)v_i) t_i(x)\Big)\, d\mu(x) \\
&= \log \int_{\mathbb{R}^n} h(x)^{\lambda+(1-\lambda)} \exp\Big(\sum_{i=1}^k \lambda u_i t_i(x)\Big)\Big(\sum_{i=1}^k (1-\lambda) v_i t_i(x)\Big)\, d\mu(x) \\
&= \log \int_{\mathbb{R}^n} \Big[h(x)\exp\Big(\sum_{i=1}^k u_i t_i(x)\Big)\Big]^\lambda \Big[h(x)\exp\Big(\sum_{i=1}^k v_i t_i(x)\Big)\Big]^{1-\lambda}\, d\mu(x) \\
[\text{Hölder}] &\leqslant \log \Big(\int_{\mathbb{R}^n} ([\ldots]^\lambda)^{1/\lambda}\, d\mu(x)\Big)^\lambda \Big(\int_{\mathbb{R}^n} ([\ldots]^{1-\lambda})^{1/(1-\lambda)}\, d\mu(x)\Big)^{1-\lambda} \\
&= \lambda \log \int_{\mathbb{R}^n} h(x)\exp\Big(\sum_{i=1}^k u_i t_i(x)\Big)\, d\mu(x) + (1-\lambda)\log\int_{\mathbb{R}^n} h(x)\exp\Big(\sum_{i=1}^k v_i t_i(x)\Big)\, d\mu(x). \qquad \square
\end{aligned}$$

For convexity of W, if $w_1, w_2 \in W$, and $\lambda \in (0,1)$, then

$$a(\lambda w_1 + (1-\lambda)w_2) \leqslant \lambda a(w_1) + (1-\lambda)a(w_2) \leqslant a(w_1) + a(w_2) < \infty.$$

This shows W is convex.

Problem 4

Using a two parameter exponential family for a Gaussian random variable (with mean μ and variance σ^2), compute both sides of the following identity in terms of μ and σ:

$$e^{-a(w)}\frac{\partial^2 e^{a(w)}}{\partial w_i \partial w_j} = \int_{\mathbb{R}} t_i(x) t_j(x) h(x) \exp\Big(\sum_{i=1}^2 w_i t_i(x) - a(w)\Big)\, d\mu(x)$$

where $1 \leqslant i, j \leqslant 2$,

$$t_1(x) := x, \qquad t_2(x) := x^2, \qquad w_1 := \frac{\mu}{\sigma^2}, \qquad w_2 := -\frac{1}{2\sigma^2},$$

and

$$a(w) := -\frac{w_1^2}{4w_2} - \frac{\log(-2w_2)}{2}.$$

Solution. For the case $i = j = 1$:

$$\begin{aligned}
e^{-a(w)}\frac{\partial^2}{\partial w_1^2} e^{a(w)} &= e^{-a(w)}\frac{\partial}{\partial w_1}\Big[\frac{\partial a(w)}{\partial w_1} e^{a(w)}\Big] \\
&= \frac{\partial^2 a(w)}{\partial w_1^2} + \Big(\frac{\partial a(w)}{\partial w_1}\Big)^2 = -\frac{1}{2w_2} + \frac{w_1^2}{4w_2^2} \\
&= \sigma^2 + \frac{\mu^2/\sigma^4}{1/\sigma^4} = \sigma^2 + \mu^2
\end{aligned}$$

and

$$\int_{\mathbb{R}} t_1^2(x) h(x) \exp(\ldots)\, d\mu(x) = \int_{\mathbb{R}} x^2 f_{\mathcal{N}(\mu,\sigma)}(x)\, dx = \mathbb{E}X^2.$$

If $i = j = 2$, then

$$\begin{aligned} e^{-a(w)}\frac{\partial^2}{\partial w_2^2}e^{a(w)} &= e^{-a(w)}\frac{\partial}{\partial w_2}\left[\frac{\partial a(w)}{\partial w_2}e^{a(w)}\right] = \frac{\partial^2 a(w)}{\partial w_2^2} + \left(\frac{\partial a(w)}{\partial w_2}\right)^2 \\ &= -\frac{w_1^2}{2w_2^3} + \frac{1}{2w_2^2} + \left(\frac{w_1^2}{4w_2^2} - \frac{1}{2w_1}\right)^2 = \frac{\mu^2/\sigma^4}{1/(4\sigma^6)} + 2\sigma^4 + \left(\frac{\mu^2/\sigma^2}{1/\sigma^2} + \sigma^2\right)^2 \\ &= 4\mu^2\sigma^2 + 2\sigma^4 + (\mu^2+\sigma^2)^2 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \end{aligned}$$

and

$$\int_{\mathbb{R}} t_2^2(x)h(x)\exp(\dots)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}} x^4 f_{\mathcal{N}(\mu,\sigma)}(x)\,\mathrm{d}x = \mathbb{E}X^4.$$

Finally, if $i \neq j$, WLOG assume $i = 1, j = 2$. Then

$$\begin{aligned} e^{-a(w)}\frac{\partial^2}{\partial w_1 \partial w_2}e^{a(w)} &= e^{-a(w)}\frac{\partial}{\partial w_1}\left[\frac{\partial a(w)}{\partial w_2}e^{a(w)}\right] = \frac{\partial^2 a(w)}{\partial w_1\partial w_2} + \frac{\partial a(w)}{\partial w_1}\frac{\partial a(w)}{\partial w_2} \\ &= \frac{w_1}{2w_2^2} - \frac{w_1}{2w_2}\left(\frac{w_1^2}{4w_2^2} - \frac{1}{2w_2}\right) \\ &= \frac{\mu/\sigma^2}{1/2\sigma^4} + \frac{\mu/\sigma^2}{1/\sigma^2}(\mu^2+\sigma^2) = \mu^3 + 3\mu\sigma^2, \end{aligned}$$

and

$$\int_{\mathbb{R}} t_1(x)t_2(x)\exp(\dots)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}} x^3 f_{\mathcal{N}(\mu,\sigma)(x)}\,\mathrm{d}x = \mathbb{E}X^3.$$

Problem 5

Let $X : \Omega \to \mathbb{R}^n$ be a random variable with

$$\mathbb{P}(X \in A) := \int_A \exp\left(-\sum_{i=1}^n x_i^2/2\right)\mathrm{d}x(2\pi)^{-n/2} \qquad \text{for all } A \subset \mathbb{R}^n \text{ measurable.}$$

Let $v \in \mathbb{R}^n$. Show that $\langle X, v\rangle \sim \mathcal{N}(0, \|v\|^2)$.

Next, let $v_1, \dots, v_m \in \mathbb{R}^n$. Show that the random variables $\langle X, v_i\rangle$ are independent if and only if the vectors $v_1, \dots, v_n$ are pairwise orthogonal.

Proof. We first show the first claim. Since X is a multivariate standard Gaussian, component-wise, each component independently follows a standard Gaussian. Thus, for a fixed v, $\langle X, v\rangle$ is merely a linear combination of Gaussians and is therefore a Gaussian. Notation-wise, let $v = (v^{(1)}, \dots, v^{(n)})$. We have

$$\mathbb{E}\langle X, v\rangle = \sum_{i=1}^n \mathbb{E}(v^{(i)}\mathcal{N}(0,1)) = 0$$

and

$$\operatorname{var}\langle X, v\rangle = \sum_{i=1}^n \operatorname{var}(v^{(i)}\mathcal{N}(0,1)) = \sum_{i=1}^n (v^{(i)})^2 = \|v\|^2.$$

For the second part, let Y_i be the random variable denoting $\langle X, v_i\rangle$ and let $Y := (Y_1, \dots, Y_n)$ be a random vector. We first prove that

> $Y_1, ..., Y_n$ are independent if and only if the moment generating functions (MGFs) satisfy $M_Y(t) = \prod_{i=1}^n M_{Y_i}(t_i)$.

The $\Rightarrow$ is obvious, as

$$M_Y(t) = \mathbb{E}e^{t^TY} = \int_{\mathbb{R}^n} \exp\Big(\sum_{i=1}^n t_i y_i\Big) \prod_{i=1}^n f_{Y_i}(y_i)\, dy$$
$$= \prod_{i=1}^n \int_{\mathbb{R}} \exp(t_i y_i) f_{Y_i}(y_i)\, dy_i = \prod_{i=1}^n \mathbb{E}e^{t_iY_i} = \prod_{i=1}^n M_Y(t_i).$$

Conversely, if $M_Y(t) = \prod_{i=1}^n M_Y(t_i)$, the result follows from the (nontrivial) fact that joint MGF uniquely gives joint distribution which, in turn, gives independence. END OF PROOF OF CLAIM.

With this claim, it suffices to prove that

> $M_Y(t) = \prod_{i=1}^n M_{Y_i}(t_i)$ if and only if $v_1, ..., v_n$ are pairwise orthogonal.

WLOG we can assume each v_i has been normalized. Let $t \in \mathbb{R}^n$ be given. Let Σ be the covariance matrix corresponding to Y. From the previous part, Y has mean 0. Thus $Y \sim \mathcal{N}_n(0, \Sigma)$ and $M_Y(t) = \exp(t^T\Sigma t/2)$. Expanding terms in Σ, we see

$$\Sigma_{i,j} = \operatorname{cov}(\langle X, v_i\rangle, \langle X, v_j\rangle) = \operatorname{cov}\Big(\sum_{k=1}^n X_k v_i^k, \sum_{\ell=1}^n X_\ell v_j^\ell\Big) = \sum_{k,\ell=1}^n v_i^k v_j^\ell \operatorname{cov}(X_k, X_\ell) = \langle v_i, v_j\rangle.$$

Therefore,

$$t^T\Sigma t = \sum_{i=1}^n t_i^2 \|v_i\|^2 + \sum_{i\neq j} t_i t_j \langle v_i, v_j\rangle. \tag{1}$$

If Y_i's are independent, the in particular they are pairwise independent, so for any $i \neq j$, if we let Σ_0 denote the covariance matrix of $Y_0 := (Y_i, Y_j)$, and let $t_0 := (t_i, t_j)$ (i.e., we only consider the corresponding two components), then from (1)

$$\frac{t_0^T\Sigma_0 t_0}{2} = \frac{t_i^2}{2} + \frac{t_j^2}{2} + t_i^2 t_j^2 \langle v_i, v_j\rangle.$$

By independence, $M_{Y_0}(t_0) = M_{Y_i}(t_i)M_{Y_j}(t_j)$, so $t_i^2t_j^2\langle u_i, v_j\rangle = 0$ for all t_i, v_j, which implies u_i, v_j are orthogonal. Ranging over all i, j's, we have that $v_1, ..., v_m$ are pairwise orthogonal.
Conversely, if $v_1, ..., v_m$ are pairwise orthogonal, then (1) implies

$$t^T\Sigma t = \sum_{i=1}^n t_i^2\|v_i\|^2 + \sum_{i\neq j} t_i t_j \cdot 0 = \sum_{i=1}^n t_i^2,$$

so $M_Y(t) = \prod_{i=1}^n M_{Y_i}(t_i)$, as expected. This finishes the proof. □

Problem 6

Show that a gamma distribution is a 2-parameter exponential family.

Then, verify its mean and variance by differentiating the exponential family.

Finally, find the moment generating function of a gamma distributed random variable and use it to find the distribution of $\sum_{i=1}^n X_i$ where $X_1, \ldots, X_n$ are independent with parameters α_i and β.

Proof. Recall that the Gamma distribution is given by

$$f(x) = \frac{x^{\alpha-1}\exp(-x/\beta)}{\beta^\alpha \Gamma(\alpha)} \qquad \text{for } x \geqslant 0.$$

Therefore,

$$\begin{aligned} f(x) &= 1_{\{x>0\}} \exp\left((\alpha-1)\log x - \frac{x}{\beta} - \alpha \log \beta - \log \Gamma(\alpha)\right) \\ &= 1_{\{x>0\}} \exp\left(-\frac{1}{b}\cdot x + (\alpha-1)\cdot \log x - (\alpha \log \beta + \log(\Gamma(\alpha)))\right) \\ &= h(x) \exp\left(\sum_{i=1}^2 w_i(\alpha,\beta)t_i(x) - a(w(\alpha,\beta))\right) \end{aligned}$$

where

$$h(x) = 1_{\{x>0\}} \qquad \begin{aligned} w_1(\alpha,\beta) &= -1/\beta \\ w_2(\alpha,\beta) &= \alpha - 1 \end{aligned} \qquad \begin{aligned} t_1(x) &= x \\ t_2(x) &= \log x \end{aligned} \qquad a(w(\alpha,\beta)) = \alpha \log \beta + \log \Gamma(\alpha). \tag{1}$$

We now verify the mean and variance using exponential family. Using the non-canonical form differential identity on β, we have

$$e^{-a(w(\alpha,\beta))}\frac{\partial}{\partial \beta} e^{a(w(\alpha,\beta))} = \mathbb{E}_{\alpha,\beta}\left(\sum_{i=1}^2 \frac{\partial w_i}{\partial \beta} t_i\right).$$

The RHS is simply $\mathbb{E}_{\alpha,\beta}(x/\beta^2) = \mathbb{E}_{\alpha,\beta}X/\beta^2$. The LHS is

$$e^{-a(w(\alpha,\beta))}e^{a(w(\alpha,\beta))}\frac{\partial}{\partial\beta}[a(w(\alpha,\beta))] = \alpha/\beta.$$

Therefore we have

$$\mathbb{E}_{\alpha,\beta}X = \beta^2 \cdot \alpha/\beta = \alpha\beta, \tag{Expected Value}$$

as expected. We now compute the second moment by differentiating twice. First note that

$$\begin{aligned} \exp(-a(w))\frac{\partial^2}{\partial w_i^2}\exp(a(w)) &= \exp(-a(w))\frac{\partial}{\partial w_i}\left(\frac{\partial}{\partial w_i}\exp(a(w))\right) \\ &= \exp(-a(w))\frac{\partial}{\partial w_i}\left(a(w)\frac{\partial}{\partial w_i}a(w)\right) \\ &= \exp(-a(w))\frac{\partial}{\partial w_i}\left(a(w)\mathbb{E}t_i\right) \\ &= \mathbb{E}t_i \exp(-a(w))\frac{\partial}{\partial w_i}a(w) = (\mathbb{E}t_i)^2 \end{aligned}$$

where we have used the canonical differential identity $\exp(-a(w))\dfrac{\partial}{\partial w_i}\exp(a(w)) = \mathbb{E}t_i$ twice. Then

$$\begin{aligned}
e^{-a(w(\alpha,\beta))}\frac{\partial^2}{\partial\beta^2}e^{a(w(\alpha,\beta))} &= e^{-a(w(\alpha,\beta))}\frac{\partial}{\partial\beta}\left[\frac{\partial}{\partial\beta}e^{a(w(\alpha,\beta))}\right] \\
&= e^{-a(w(\alpha,\beta))}\frac{\partial}{\partial\beta}\left[\frac{\partial e^{a(w)}}{\partial w_1}\frac{\partial w_1}{\partial\beta} + \frac{\partial e^{a(w)}}{\partial w_2}\frac{\partial w_2}{\partial\beta}\right] \\
&= e^{-a(w(\alpha,\beta))}\left[\frac{\partial^2 e^{a(w)}}{\partial w_1^2}\left(\frac{\partial w_1}{\partial\beta}\right)^2 + \frac{\partial e^{a(w)}}{\partial w_1}\frac{\partial^2 w_1}{\partial^2\beta} + \frac{\partial^2 e^{a(w)}}{\partial w_2^2}\left(\frac{\partial w_2}{\partial\beta}\right)^2 + \frac{\partial e^{a(w)}}{\partial w_2}\frac{\partial^2 w_2}{\partial\beta}\right] \\
&= \mathbb{E}_{\alpha,\beta}\left(\sum_{i=1}^2\left(t_i\cdot\frac{\partial w_i}{\partial\beta}\right)^2 + \sum_{i=1}^2 t_i\cdot\frac{\partial^2 w_i}{\partial\beta^2}\right) \\
&= \mathbb{E}_{\alpha,\beta}X^2/\beta^4 - 2\mathbb{E}_{\alpha,\beta}X/\beta^3 = \mathbb{E}_{\alpha,\beta}X^2/\beta^4 - 2\alpha/\beta^2. \qquad (2)
\end{aligned}$$

The original term, on the other hand, also satisfies

$$\begin{aligned}
e^{-a(w(\alpha,\beta))}\frac{\partial^2}{\partial\beta^2}e^{a(w(\alpha,\beta))} &= e^{-a(w(\alpha,\beta))}\frac{\partial}{\partial\beta}\left(e^{a(w(\alpha,\beta))}\frac{\partial}{\partial\beta}a(w(\alpha,\beta))\right) \\
&= e^{-a(w(\alpha,\beta))}\left(e^{a(w(\alpha,\beta))}\frac{\partial^2}{\partial\beta^2}a(w(\alpha,\beta)) + e^{a(w(\alpha,\beta))}\left(\frac{\partial}{\partial\beta}a(w(\alpha,\beta))\right)^2\right) \\
&= \frac{\partial^2}{\partial\beta^2}a(w(\alpha,\beta)) + \left(\frac{\partial}{\partial\beta}a(w(\alpha,\beta))\right)^2 = -\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}. \qquad (3)
\end{aligned}$$

Combining (2) and (3) we see

$$\mathbb{E}_{\alpha,\beta}X^2 = \beta^4\beta^2(\alpha + \alpha^2 + 2\alpha) = \alpha\beta^2 + \alpha^2\beta^2,$$

so

$$\text{var}_{\alpha,\beta}(X) = \alpha\beta^2 + \alpha^2\beta^2 - \alpha^2\beta^2 = \alpha\beta^2, \qquad \text{(Variance)}$$

also as expected.

Finally we compute the MGF $M_{X,\alpha,\beta}(t)$ for $t < \beta^{-1}$:

$$\begin{aligned}
M_{X,\alpha,\beta}(t) = \mathbb{E}_{\alpha,\beta}e^{tX} &= \int_0^\infty \frac{e^{tx}x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\,\mathrm{d}x = \int_0^\infty \frac{x^{\alpha-1}e^{-x(\beta^{-1}-t)}}{\beta^\alpha\Gamma(\alpha)}\,\mathrm{d}x \\
&= \int_0^\infty \frac{(u/(\beta^{-1}-t))^{\alpha-1}e^{-u}}{\beta^\alpha\Gamma(\alpha)}\frac{1}{\beta^{-1}-t}\,\mathrm{d}u \\
&= \frac{1}{(\beta^{-1}-t)^\alpha\beta^\alpha\Gamma(\alpha)}\int_0^\infty u^{\alpha-1}e^{-u}\,\mathrm{d}u = \frac{1}{(\beta^{-1}-t)^\alpha\beta^\alpha} = (1-\beta t)^{-\alpha}. \qquad \text{(MGF)}
\end{aligned}$$

If $X_1,\dots,X_n$ are independent with $X_i \sim \text{Gamma}(\alpha_i,\beta)$, then the MGF at t becomes

$$\prod_{i=1}^n (1-\beta t)^{\alpha_i} = (1-\beta t)^{-\sum_{i=1}^n \alpha_i}.$$

By uniqueness, this implies that $\sum_{i=1}^n X_i$ is a $\left(\sum_{i=1}^n \alpha_i, \beta\right)$–distributed Gamma random variable. □

Problem 7

† Let $n \geqslant 2$ be an integer. Let $X_1, ..., X_n$ be a random sample of size n (that is, $X_1, ..., X_n$ are i.i.d. random variables). Assume that $\mu := \mathbb{E}X_1 \in \mathbb{R}$ and $\sigma := \sqrt{\operatorname{var}(X_1)} < \infty$. Let $\overline{X}$ be the sample mean and let S be the sample standard deviation of the random sample. Show that $\operatorname{var}(\overline{X}) = \sigma^2/n$ and $\mathbb{E}S^2 = \sigma^2$.

Proof. By definition we have $\overline{X} = (X_1 + ... + X_n)/n$ and $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$. Then,

$$\operatorname{var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

To show that $\mathbb{E}S^2 = \sigma^2$,

$$\begin{aligned}
\mathbb{E}S^2 &= \frac{1}{n-1}\mathbb{E}\sum_{i=1}^n (X_i - \overline{X})^2 \\
&= \frac{1}{n-1}\mathbb{E}\sum_{i=1}^n (X_i^2 - 2X_i\overline{X} + \overline{X}^2) \\
&= \frac{1}{n-1}\left[n\mathbb{E}X_1^2 - 2\mathbb{E}\Big(\overline{X}\sum_{i=1}^n X_i\Big) + n\mathbb{E}\overline{X}^2\right] \\
&= \frac{1}{n-1}\left[n\mathbb{E}X_1^2 - 2\mathbb{E}n\overline{X}^2 + n\mathbb{E}\overline{X}^2\right] \\
&= \frac{1}{n-1}\left[n\mathbb{E}X_1^2 - n\mathbb{E}\overline{X}^2\right]. \qquad (1)
\end{aligned}$$

Since

$$\operatorname{var}(X_1) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2,$$

we obtain

$$\mathbb{E}X_1^2 = \sigma^2 + \mu^2 \qquad \text{and likewise} \qquad \mathbb{E}\overline{X} = \operatorname{var}(\overline{X}) + \mu^2 = \frac{\sigma^2}{n} + \mu^2.$$

Substituting these values back into (1), we obtain $\mathbb{E}S^2 = (n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2)/(n-1) = \sigma^2$, as claimed. □

Problem 8

Let $X : \Omega \to \mathbb{R}$ be a random variable with $\mathbb{E}X^2 < \infty$. Show that $\mathbb{E}(X - t)^2$ is uniquely minimized when $t = \mathbb{E}X$.

Proof. By linearity,

$$\mathbb{E}(X - t)^2 = \mathbb{E}X^2 - 2t\mathbb{E}X + \mathbb{E}t^2 = \mathbb{E}X^2 - 2t\mathbb{E}X + t^2.$$

This is the sum of a constant, a linear, and a strictly convex function. Therefore it is strictly convex. Since $t = \mathbb{E}X$ is a critical point with second derivative 2, it is the unique global minimum. □

Problem 9

† Let X be a chi squared random variables with p degrees of freedom. Let Y be a chi squared random variable with q degrees of freedom. Assume that X and Y are independent. Show that $(X/p)/(Y/q)$ has the following density, known as the **Snedecor's f-distribution** with p and q degrees of freedom

$$f_{(X/p)/(Y/q)}(t) := \frac{t^{p/2-1}(p/q)^{p/2}\Gamma((p+q)/2)}{\Gamma(p/2)\Gamma(q/2)}(1+t(p/q))^{-(p+q)/2} \qquad \text{for all } t>0.$$

Proof. Let X and Y be as stated. By definition, we have the PDFs

$$f_X(x) = \frac{x^{p/2-1}e^{-x/2}}{2^{p/2}\Gamma(p/2)} \qquad \text{and} \qquad f_Y(y) = \frac{y^{q/2-1}e^{-y/2}}{2^{q/2}\Gamma(q/2)}. \tag{1}$$

By independence, we also have the JPDF

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{x^{p/2-1}y^{q/2-1}e^{-(x+y)/2}}{2^{(p+q)/2}\Gamma(p/2)\Gamma(q/2)}. \tag{2}$$

Note that $(X/p)/(Y/q) = (X/Y)(q/p)$ and q/p is a constant, so the important part is to compute X/Y. We begin by computing its CDF. Let $t>0$. Then

$$\begin{aligned} F_{X/Y}(t) &= P(X/Y \leqslant t) = P(X \leqslant tY) \\ &= \int_0^\infty \int_0^{yt} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2^{(p+q)/2}\Gamma(p/2)\Gamma(q/2)} \int_0^\infty \left[\int_0^{yt} x^{p/2-1}e^{-x/2}\,\mathrm{d}x\right] y^{q/2-1}e^{-y/2}\,\mathrm{d}y. \end{aligned} \tag{3}$$

We can recover the PDF of X/Y by differentiating (3) with respect to t:

$$\begin{aligned} f_{X/Y}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(3) &= \frac{1}{2^{(p+q)/2}\Gamma(p/2)\Gamma(q/2)} \int_0^\infty \left[(yt)^{p/2-1}e^{-yt/2}\cdot y\right] y^{q/2-1}e^{-y/2}\,\mathrm{d}y \\ &= \frac{t^{p/2-1}}{2^{(p+q)/2}\Gamma(p/2)\Gamma(q/2)} \int_0^\infty y^{(p+q)/2-1}\cdot e^{-y(t+1)/2}\,\mathrm{d}y \\ (\Delta) &= \frac{t^{p/2-1}}{2^{(p+q)/2}\Gamma(p/2)\Gamma(q/2)}\cdot \Gamma(p/2+q/2)\left(\frac{2}{t+1}\right)^{(p+q)/2} \\ &= \frac{\Gamma((p+q)/2)}{\Gamma(p/2)\Gamma(q/2)}\cdot\frac{t^{p/2-1}}{(t+1)^{(p+q)/2}}, \end{aligned} \tag{4}$$

where (Δ) is because

$$g(y) := \frac{y^{(p+q)/2-1}\cdot e^{-y(t+1)/2}}{(2/(t+1))^{(p+q)/2}\cdot\Gamma((p+q)/2}$$

is the PDF of a $\left(\frac{p+q}{2}, \frac{2}{t+1}\right)$–distributed Gamma random variable and thus has integral 1. Finally,

$$\begin{aligned} f_{(X/p)(Y/q)}(t) &= f_{(X/Y)(q/p)}(t) = \frac{p}{q}\cdot f_{X/Y}(t(p/q)) \\ [\text{by } (4)] &= \frac{p}{q}\cdot\frac{\Gamma((p+q)/2)}{\Gamma(p/2)\Gamma(q/2)}\cdot\frac{(t(p/q))^{p/2-1}}{(1+t(p/q))^{(p+q)/2}} \\ &= \frac{t^{p/2-1}(p/q)^{p/2}}{(1+t(p/q))^{(p+q)/2}}\cdot\frac{\Gamma((p+q)/2)}{\Gamma(p/2)\Gamma(q/2)}, \end{aligned}$$

as claimed. □