Solutions to Problem Sheet 3: Advanced Bayesian Statistics

1. **Conditional Distributions**
   (a) Give an example of two different bivariate distributions admitting the same marginal distributions.

   Solution: Consider two variables X, Y such that $X \sim \mathcal{N}(0,1)$ and $Y = WX$ where W is a discrete random variable independent of X with $P(W=1) = P(W=-1) = 1/2$. While the marginals are both Gaussian, the pair (X, Y) is not bivariate Gaussian, even though the marginals are the same as those of $\mathcal{N}(0, I_2)$. To see that the pair is not Gaussian, remark that $P(X+Y=0) = P(W=-1) = 1/2$ which contradicts the fact that $X+Y$ should be a continuous random variables or a single point-mass.

   Alternatively, one can use Copulas to construct different joint dependencies for the same marginals (uniform distributions on $[0,1]$ in particular).

   (b) Show that knowledge of the conditional distributions $X|Y$ and $Y|X$ is sufficient to characterize the distribution of (X, Y).

   Solution: In the density framework, we have $f(x,y) = f(x|y)f_Y(y) = f(y|x)f_X(x)$. This implies:
$$\frac{f_X(x)}{f_Y(y)} = \frac{f(x|y)}{f(y|x)}$$
   By integrating both sides with respect to x, we can isolate the marginals and thus reconstruct the full joint density $f(x,y)$. In particular,
$$f_Y(y) = \left[\int \frac{f(x|y)}{f(y|x)}dx\right]^{-1}, \quad f_X(x) = \frac{f(x|y)}{f(y|x)}\left[\int \frac{f(x|y)}{f(y|x)}dx\right]^{-1}.$$

2. **Risk Types (Review)**
   (a) Recall the definitions of:
      i. The risk function: $R(\theta, T) = \mathbb{E}_\theta[L(\theta, T(X))]$.
      ii. The Bayes risk: $R_B(\Pi, T) = \mathbb{E}_{\theta\sim\Pi}[R(\theta, T)]$.

   (b) Recall the definitions:

   Solution:

   - **Admissible**: An estimator T is admissible if no other estimator T' satisfies $R(\theta, T') \leq R(\theta, T)$ for all θ with strict inequality for at least one θ.

- **Bayes estimator**: $T_\Pi = \arg\min_T R_B(\Pi, T)$.
- **Minimax**: T^* is minimax if $\sup_\theta R(\theta, T^*) = \inf_T \sup_\theta R(\theta, T)$.

3. **Mixtures**

(a) Show that a random variable Y with distribution Q can be obtained by the scheme $Z \sim \mathcal{B}ern(\rho)$, $Y|Z \sim Q_Z$.

Solution: Let g be any bounded measurable function. By the law of total expectation (conditioning on Z):

$$\begin{aligned}\mathbb{E}[g(Y)] &= \mathbb{E}[\mathbb{E}[g(Y)|Z]] \\ &= P(Z=0)\int g(y)dQ_0(y) + P(Z=1)\int g(y)dQ_1(y) \\ &= (1-\rho)\int g(y)q_0(y)d\mu(y) + \rho\int g(y)q_1(y)d\mu(y) \\ &= \int g(y)\left[(1-\rho)q_0(y) + \rho q_1(y)\right]d\mu(y).\end{aligned}$$

Since this holds for any g, the density of Y is $(1-\rho)q_0 + \rho q_1$, which means $Y \sim Q$.

(b) Gaussian Mixture Model.

i. Determine the density of the posterior distribution $\theta|X$.

Solution: Let $\phi_{\mu,\sigma^2}(\theta)$ denote the density of $\mathcal{N}(\mu, \sigma^2)$. The prior is $\pi(\theta) = (1-\rho)\phi_{0,1}(\theta) + \rho\phi_{5,1}(\theta)$. Using Bayes' formula, the posterior density is proportional to Likelihood $\times$ Prior:

$$f(\theta|X) \propto e^{-\frac{1}{2}\sum_{i=1}^n (X_i-\theta)^2}\left[(1-\rho)e^{-\frac{\theta^2}{2}} + \rho e^{-\frac{(\theta-5)^2}{2}}\right].$$

ii. Show that the posterior is a two-component mixture and find the weight $\rho_n(X)$.

Solution: We can rewrite the product of the likelihood and the prior components by completing the square.

- **Component 0:** Prior $\mathcal{N}(0,1) \implies$ Posterior kernel $\mathcal{N}(\frac{n\bar{X}}{n+1}, \frac{1}{n+1})$.
- **Component 1:** Prior $\mathcal{N}(5,1) \implies$ Posterior kernel $\mathcal{N}(\frac{n\bar{X}+5}{n+1}, \frac{1}{n+1})$.

Thus, the posterior density is a mixture:

$$f(\theta|X) = (1-\rho_n(X))\phi_{\frac{n\bar{X}}{n+1},\frac{1}{n+1}}(\theta) + \rho_n(X)\phi_{\frac{n\bar{X}+5}{n+1},\frac{1}{n+1}}(\theta).$$

To find the weight $\rho_n(X)$, we define the marginal likelihoods (evidence) under each component prior as $D_X^{(0)}$ and $D_X^{(1)}$:

$$\rho_n(X) = \frac{\rho D_X^{(1)}}{(1-\rho)D_X^{(0)} + \rho D_X^{(1)}} = \frac{\rho}{(1-\rho)\frac{D_X^{(0)}}{D_X^{(1)}} + \rho}.$$

(c) Coin toss (Beta Mixture Prior).

i. Justify the prior.

Solution: The Beta distribution is conjugate to the Bernoulli likelihood. Also, the Beta component with $\alpha < \beta$ assigns more mass to $\theta > 1/2$.

ii. Determine the posterior and weights.

Solution: Let $S_n = \sum_{i=1}^n X_i$. Since the Beta distribution is conjugate to the Binomial likelihood:

- Component 0: Prior Beta$(2, 4) \to$ Posterior Beta$(2 + S_n, 4 + n - S_n)$.
- Component 1: Prior Beta$(3, 3) \to$ Posterior Beta$(3 + S_n, 3 + n - S_n)$.

The posterior is a mixture of these two Beta distributions with weight $\rho_n(X)$. Using the formula from part (c), the weight depends on the ratio of marginal likelihoods $D_X^{(0)}/D_X^{(1)}$.

$$\frac{D_X^{(0)}}{D_X^{(1)}} = \frac{B(3,3)}{B(2,4)} \cdot \frac{B(2+S_n, 4+n-S_n)}{B(3+S_n, 3+n-S_n)}.$$

1. Ratio of Prior Constants:

$$\frac{B(3,3)}{B(2,4)} = \frac{2}{3}.$$

2. Ratio of Posterior Constants:

$$\frac{\Gamma(S_n+2)\Gamma(n-S_n+4)}{\Gamma(n+6)} \Bigg/ \frac{\Gamma(S_n+3)\Gamma(n-S_n+3)}{\Gamma(n+6)} = \frac{n-S_n+3}{S_n+2}.$$

Combining these:

$$\frac{D_X^{(0)}}{D_X^{(1)}} = \frac{2}{3}\frac{n-S_n+3}{S_n+2}.$$

This ratio is plugged into the weight formula:

$$\rho_n(X) = \frac{\rho}{(1-\rho)\frac{D_X^{(0)}}{D_X^{(1)}} + \rho}.$$

(d) General Phenomenon.

Solution: If the prior is a mixture $\pi(\theta) = \sum \rho_k \pi_k(\theta)$, and $X|\theta \sim p_\theta$, the posterior is always a mixture of the component posteriors $\pi_k(\theta|X)$.
Let $D_X^{(k)} = \int p_\theta(X)\pi_k(\theta)d\theta$ be the marginal likelihood (evidence) under component k. The posterior density is:

$$f(\theta|X) = \sum_k \rho_{n,k}(X) f(\theta|X,k)$$

where the updated weights are proportional to the prior weights times the evidence:

$$\rho_{n,k}(X) \propto \rho_k D_X^{(k)}.$$

Specifically for two components:

$$\rho_n(X) = \frac{\rho D_X^{(1)}}{(1-\rho)D_X^{(0)} + \rho D_X^{(1)}}.$$

4. **Bayes Estimators**

Let $X|\theta \sim \mathcal{N}(\theta, 1)$ and $\theta \sim \mathcal{N}(0, \sigma^2)$.

(a) Give the Bayes estimator of θ for the quadratic loss function $L(\theta, T) = (\theta - T)^2$.

Solution: The Bayes estimator for the quadratic loss is the mean of the posterior distribution. The posterior distribution is $\theta|X \sim \mathcal{N}\left(\frac{X}{1+\sigma^{-2}}, \frac{1}{1+\sigma^{-2}}\right)$ (using the standard formula for Gaussian conjugacy).
Thus, the Bayes estimator is:

$$\hat{\theta}_{Bayes} = \mathbb{E}[\theta|X] = \frac{X}{1+\sigma^{-2}} = \frac{\sigma^2 X}{1+\sigma^2}.$$

(b) Assume $\sigma^2 \leq 1$. Consider the weighted loss function $L_w(\theta, T) = \exp\{\frac{3\theta^2}{4}\}(\theta - T)^2$. Determine the associated Bayes estimator and its Bayes risk.

Solution:
1. Determining the Estimator T^*
A Bayes estimator minimizes the posterior expected loss. We seek T that minimizes:

$$\int L_w(\theta, T) d\Pi(\theta|X) = \int (\theta - T)^2 e^{\frac{3\theta^2}{4}} \pi(\theta|X) d\theta.$$

Using the posterior density $\pi(\theta|X) \propto \exp\left\{-\frac{1+\sigma^{-2}}{2}\left(\theta - \frac{X}{1+\sigma^{-2}}\right)^2\right\}$, the integrand is proportional to:

$$(\theta - T)^2 \exp\left\{-\left(\frac{1+\sigma^{-2}}{2} - \frac{3}{4}\right)\theta^2 + \theta X\right\}.$$

We simplify the coefficient of θ^2 in the exponent:

$$\frac{1+\sigma^{-2}}{2} - \frac{3}{4} = \frac{2+2\sigma^{-2}-3}{4} = \frac{2\sigma^{-2}-1}{4} = \frac{\sigma^{-2}-1/2}{2}.$$

We recognize the kernel of a new Gaussian distribution Q with variance $1/(\sigma^{-2}-1/2)$ and mean:

$$\mu_Q = \frac{X}{\sigma^{-2}-1/2}.$$

The problem is equivalent to minimizing $\mathbb{E}_Q[(\theta - T)^2]$, which is minimized at the mean of Q. Thus, the Bayes estimator is:

$$T^*(X) = \frac{X}{\sigma^{-2}-1/2}.$$

2. Calculating the Bayes Risk

The Bayes risk is defined as $R_B(\Pi, T^*) = \int R(\theta, T^*(X))d\Pi(\theta)$.

$$R_B(\Pi, T^*) = \iint e^{\frac{3\theta^2}{4}} \left(\theta - \frac{x}{\sigma^{-2}-1/2}\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} dx d\Pi(\theta).$$

First, we compute the inner integral with respect to x (expectation over $X|\theta \sim \mathcal{N}(\theta, 1)$). Let $K = \sigma^{-2} - 1/2$. Then $T^*(X) = X/K$.

$$\mathbb{E}_{X|\theta}\left[\left(\theta - \frac{X}{K}\right)^2\right] = \frac{1}{K^2}\mathbb{E}_{X|\theta}[(K\theta - X)^2].$$

Since $X = \theta + \epsilon$ with $\epsilon \sim \mathcal{N}(0,1)$:

$$K\theta - X = K\theta - (\theta + \epsilon) = (K-1)\theta - \epsilon.$$

$$\mathbb{E}[((K-1)\theta - \epsilon)^2] = (K-1)^2\theta^2 + \mathbb{E}[\epsilon^2] = (K-1)^2\theta^2 + 1.$$

Substituting $K = \sigma^{-2} - 0.5$, we have $K - 1 = \sigma^{-2} - 1.5$. The inner integral result is:

$$\left(\frac{\sigma^{-2}-3/2}{\sigma^{-2}-1/2}\right)^2 \theta^2 + \left(\frac{1}{\sigma^{-2}-1/2}\right)^2.$$

Now, we integrate with respect to the prior $\theta \sim \mathcal{N}(0, \sigma^2)$. The integrand includes the weight $e^{3\theta^2/4}$.

$$R_B = \frac{1}{\sqrt{2\pi}\sigma} \int \left[A\theta^2 + B\right] e^{\frac{3\theta^2}{4}} e^{-\frac{\theta^2}{2\sigma^2}} d\theta = \frac{C}{\sigma}\frac{1}{\sqrt{2\pi}C}\int [A\theta^2 + B] e^{-\theta^2/2C^2} d\theta.$$

We obtain $R_B = \frac{C}{\sigma}[AC^2 + B]$ where $A = \left(\frac{\sigma^{-2}-3/2}{\sigma^{-2}-1/2}\right)^2$, $B = \left(\frac{1}{\sigma^{-2}-1/2}\right)^2$ and $C^2 = \frac{1}{\sigma^{-2}-3/2}$.

Convergence Condition: The integral converges only if the coefficient in the exponent is negative:

$$\frac{1}{2\sigma^2} - \frac{3}{4} > 0 \iff \frac{1}{2\sigma^2} > \frac{3}{4} \iff \sigma^2 < \frac{2}{3}.$$

If $\sigma^2 \geq 2/3$, the risk is infinite ($+\infty$).

Assuming $\sigma^2 < 2/3$, let $c = (\sigma^{-2} - 3/2)^{-1}$. The integral simplifies to:

$$R_B(\Pi, T^*) = \frac{1}{\sigma^{-2} - 1/2} \frac{1}{\sqrt{1 - \frac{3}{2}\sigma^2}}.$$

5. **Minimaxity**

(a) Prove that if an admissible estimator has constant risk, it is minimax.

Solution: This is a standard result (cf. Lecture). *Proof idea:* Suppose T is admissible with constant risk c. If T were not minimax, there would exist an estimator T' such that $\sup_\theta R(\theta, T') < \sup_\theta R(\theta, T) = c$. This would imply $R(\theta, T') < c = R(\theta, T)$ for all θ, contradicting the admissibility of T.

(b) Let T be a Bayes estimator for Π.

i. Assume $R(\theta, T) \leq R_B(\Pi, T)$ for all $\theta \in \Theta$. Show that T is minimax.

Solution: We take the maximum of the assumed inequality over θ to obtain:

$$R_{\max}(T) = \sup_\theta R(\theta, T) \leq R_B(\Pi, T).$$

Since T is the Bayes estimator for Π by hypothesis, $R_B(\Pi, T) = R_B(\Pi)$ (the minimal Bayes risk). We know from the course (minimax theorem) that the Bayes risk for any prior Π always lower bounds the minimax risk R_M:

$$R_B(\Pi) \leq R_M.$$

Combining these, we get $R_{\max}(T) \leq R_M$. However, by definition of the minimax risk, $R_M \leq R_{\max}(T)$. Therefore, we must have equality, which proves that T is minimax.

Remark: One can actually show that the function $\theta \to R(\theta, T)$ is constant Π-almost everywhere. Since $\int [R_{\max}(T) - R(\theta, T)] d\Pi(\theta) = 0$ and the integrand is non-negative, the integrand must be zero almost surely.

ii. In particular, recover the result that a Bayes estimator of constant risk is minimax.

Solution: If T is a Bayes estimator with constant risk, say $R(\theta, T) = r$, then its Bayes risk is also r (since it integrates to r). The inequality from the previous question becomes an equality:

$$R(\theta, T) = r \implies R_B(\Pi, T) = r.$$

Thus, the condition $R(\theta, T) \leq R_B(\Pi, T)$ is satisfied, and T is minimax.

6. **Quantiles**

Let the loss function be defined as $l(\theta, T) = k_2(\theta - T)$ if $\theta > T$ and $k_1(T - \theta)$ if $\theta \leq T$.

Solution: We observe that if $k_1 = k_2 = 1$, this corresponds to the absolute error loss, for which the Bayes estimator is known to be the posterior median.

To find the Bayes estimator $T^*(X)$ generally, we seek the minimizer of the posterior expected loss:

$$T^*(X) = \underset{T(X)}{\operatorname{argmin}} \int l(\theta, T(X)) d\Pi(\theta|X).$$

Let $G(\theta)$ be the cumulative distribution function (CDF) of the posterior distribution and $g(\theta)$ be its density. We can split the integral at T:

$$\begin{aligned}
\Psi(T) &= \int_{-\infty}^{T} k_1(T - \theta)g(\theta)d\theta + \int_{T}^{\infty} k_2(\theta - T)g(\theta)d\theta \\
&= k_1 T G(T) - k_1 \int_{-\infty}^{T} \theta g(\theta)d\theta + k_2 \int_{T}^{\infty} \theta g(\theta)d\theta - k_2 T(1 - G(T)).
\end{aligned}$$

We differentiate this expression with respect to T using Leibniz's rule. Note that the boundary terms cancel out or are zero (since the integrand is zero at $\theta = T$).

$$\Psi'(T) = \int_{-\infty}^{T} k_1 g(\theta)d\theta + \int_{T}^{\infty} (-k_2)g(\theta)d\theta = k_1 G(T) - k_2(1 - G(T)).$$

Setting the derivative to zero to find the minimum:

$$k_1 G(T) - k_2(1 - G(T)) = 0 \iff (k_1 + k_2)G(T) = k_2.$$

Thus, the condition for the estimator T is:

$$G(T) = \frac{k_2}{k_1 + k_2}.$$

Conclusion: The Bayes estimator $T(X)$ is the quantile of order $\frac{k_2}{k_1+k_2}$ of the posterior distribution $\Pi[\cdot|X]$. This generalizes the median result (where $k_1 = k_2 \implies$ quantile $1/2$).

7. **The Hodges Phenomenon**

Let $T_n = \bar{X}$ and $S_n = T_n \mathbb{I}_{|T_n| \geq n^{-1/4}}$.

(a) Calculate the quadratic risk of T_n.

> **Solution:** Since $X_i \sim \mathcal{N}(\theta, 1)$, $\bar{X} \sim \mathcal{N}(\theta, 1/n)$. The quadratic risk is simply the variance:
> $$R(\theta, T_n) = \mathbb{E}_\theta[(\bar{X} - \theta)^2] = \mathrm{Var}(\bar{X}) = \frac{1}{n}.$$
> This risk is constant for all θ.

(b) Show the lower bound inequality for S_n.

> **Solution:** We lower bound the risk of S_n by considering the case where the estimator is zero (i.e., $|T_n| < n^{-1/4}$):
> $$\begin{aligned} \mathbb{E}_\theta[(S_n - \theta)^2] &= \mathbb{E}_\theta[(S_n - \theta)^2 \mathbb{I}_{|T_n| \leq n^{-1/4}}] + \mathbb{E}_\theta[(S_n - \theta)^2 \mathbb{I}_{|T_n| > n^{-1/4}}] \\ &\geq \mathbb{E}_\theta[(0 - \theta)^2 \mathbb{I}_{|T_n| \leq n^{-1/4}}] \\ &= \theta^2 P_\theta[|T_n| \leq n^{-1/4}]. \end{aligned}$$
> Since $T_n \sim \mathcal{N}(\theta, 1/n)$ under P_θ, we have the result:
> $$\mathbb{E}_\theta[(S_n - \theta)^2] \geq \theta^2 P\left[\left|\mathcal{N}\left(\theta, \frac{1}{n}\right)\right| \leq n^{-1/4}\right].$$

(c) Deduce the existence of the constant c.

> **Solution:** We verify the lower bound for a specific sequence of parameters $\theta^* = n^{-1/4}/2$. Substituting θ^* into the probability term:
> $$P\left[\left|\mathcal{N}\left(\frac{1}{2n^{1/4}}, \frac{1}{n}\right)\right| \leq \frac{1}{n^{1/4}}\right] = P\left[\left|\frac{1}{2n^{1/4}} + \frac{1}{\sqrt{n}}\mathcal{N}(0,1)\right| \leq \frac{1}{n^{1/4}}\right].$$
> The above gives
> $$P\left[\left|\frac{1}{2n^{1/4}} + \frac{1}{\sqrt{n}}Z\right| \leq \frac{1}{n^{1/4}}\right] \geq P\left[\frac{1}{2n^{1/4}} + \left|\frac{1}{\sqrt{n}}Z\right| \leq \frac{1}{n^{1/4}}\right] = P\left[|Z| \leq n^{1/4}/2\right].$$
> Indeed, the event $\left\{\frac{1}{\sqrt{n}}|Z| \leq \frac{1}{2n^{1/4}}\right\}$ implies the event $\left\{|\theta^* + \frac{1}{\sqrt{n}}Z| \leq \frac{1}{n^{1/4}}\right\}$ because $|\theta^*| = \frac{1}{2n^{1/4}}$.
> Since $n \geq 1$, $n^{1/4}/2 \geq 1/2$. So the probability is bounded by $c_0 = P[|Z| \leq 1/2] > 0$.
> The risk at θ^* is thus:
> $$\mathbb{E}_{\theta^*}[(S_n - \theta^*)^2] \geq (\theta^*)^2 c_0 = \frac{1}{4\sqrt{n}} c_0.$$
> Multiplying by n:
> $$n\mathbb{E}[(S_n - \theta)^2] \geq \frac{c_0}{4}\sqrt{n}.$$
> This proves the result with $c = c_0/4$.

(d) Compare the maximal risk. Is S_n "uniformly good"?

Solution: The maximal risk of S_n is at least of the order $n^{-1/2}$ (since nRisk $\sim \sqrt{n} \implies$ Risk $\sim 1/\sqrt{n}$). The maximal risk of T_n is exactly $1/n$.

Since $1/\sqrt{n} \gg 1/n$ for large n, the estimator S_n has a much worse worst-case scenario than T_n.

Remarks: One can show that, from the point of view of pointwise risk, we have for any fixed non-zero θ:

$$\lim_{n\to\infty} nR(\theta, S_n) = \lim_{n\to\infty} nR(\theta, T_n) = 1.$$

For $\theta = 0$, one can show that

$$\lim_{n\to\infty} nR(0, S_n) = o(1),$$

whereas $\lim_{n\to\infty} nR(0, T_n) = 1$. One might therefore think that S_n is "better" than T_n. According to the exercise, this is not the case if one considers a uniform measure of risk, such as the maximal risk.In general, it is very important in statistics to have uniform results if one wishes to avoid phenomena like the one described in the exercise, where the risk of S_n becomes very large in the neighborhood of certain points (note that these points move with n, so this is not visible if one considers a limit at a single point as $n \to \infty$, but the phenomenon will be visible if n is finite in a non-asymptotic framework).

8. **Bayesian Estimator of the Loss Function**

Solution: We seek the estimator $\hat{L}(X)$ that minimizes the Bayes risk associated with the quadratic meta-loss $\tilde{L}$.

1. Formulation of the Bayes Risk Let $Y = L(\theta, T(X))$ be the quantity we wish to estimate (the loss itself). Note that once the data X is observed, $T(X)$ is a fixed value, so Y becomes a random variable depending only on the unknown parameter θ.

The Bayes risk for the estimator $\hat{L}$ is the expectation of the meta-loss over the joint distribution of X and θ:

$$R_{meta}(\Pi, \hat{L}) = \mathbb{E}_{X,\theta}\left[\tilde{L}\right] = \mathbb{E}_{X,\theta}\left[(L(\theta, T(X)) - \hat{L}(X))^2\right].$$

2. Minimization via Posterior Expectation To find the Bayes estimator, we minimize the posterior expected loss pointwise for each observation X. We rewrite the risk by conditioning on X:

$$R_{meta}(\Pi, \hat{L}) = \mathbb{E}_X\left[\mathbb{E}_{\theta|X}\left[(L(\theta, T(X)) - \hat{L}(X))^2\right]\right].$$

Since the outer expectation $\mathbb{E}_X$ is non-negative, minimizing the total risk is equivalent to minimizing the inner term for every possible observation x:

$$\hat{L}(x) = \underset{c\in\mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}_{\theta|x}\left[(L(\theta, T(x)) - c)^2\right].$$

3. Solving the Quadratic Minimization This is the standard problem of minimizing the Mean Squared Error (MSE). For any random variable Z with finite variance, the constant c that minimizes the quantity $\mathbb{E}[(Z-c)^2]$ is the expectation $\mathbb{E}[Z]$.

Here, our "random variable" is the loss function $L(\theta, T(x))$ itself, distributed according to the posterior distribution of θ.

Conclusion Therefore, the Bayes estimator of the loss is the **posterior expected loss**:

$$\hat{L}(X) = \mathbb{E}_{\theta|X}[L(\theta, T(X))] = \int_{\Theta} L(\theta, T(X))\pi(\theta|X)d\theta.$$