Bibliography

[1] V. Abrash, A. Sankar, H. Franco, and M. Cohen. Acoustic adaptation using nonlinear transformations of HMM parameters. In *Proc. ICASSP '96*, pages 729 – 732, Atlanta, GA, May 1996.

[2] S. Amari. A theory of adaptive pattern classifiers. *IEEE Trans. Electron. Comput.*, 16:299–307, June 1967.

[3] H.C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics*, 72:2384–2393, 1980.

[4] ARPA. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Training and Test Data. NIST Speech Disc CD1-1.1, Dec. 1990.

[5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

[6] J. R. Bellegarda, P. V. de Souza, D. Nahamoo, M. Padmanabhan, M. A. Picheny, and L. R. Bahl. Experiments using data augmentation for speaker adaptation. In *Proc. ICASSP '95*, pages 692 – 695, Detroit, MI, May 1995.

[7] J.M. Bernardo and A.F.M. Smith. *Bayesian theory.* Wiley, England, 1994.

[8] A. Biem and S. Katagiri. Feature extraction based on minimum classification error/generalized probabilistic descent method. In *Proc. ICASSP '93*, pages II–275 – II–278, Minneapolis, MN, April 1993.

[9] A. Biem and S. Katagiri. Filter bank design based on discriminative feature extraction. In *Proc. ICASSP '94*, pages I–485 – I–488, Adelaide, Australia, April 1994.

[10] J. Billa, K. Ma, J.W. McDonough, G. Zavaliagkos, D.R. Miller, K.N. Ross, and A. El-Jaroudi. Multilingual speech recognition: The 1996 Byblos callhome system. In *Proc. Eurospeech '97*, pages 363–366, Rhodes, Greece, September 1997.

[11] C.M. Bishop. Curvature-driven smoothing: A learning algorithm for feed-forward networks. *IEEE Trans. Neural Networks*, 4(5):882–884, 1993.

[12] C.M. Bishop. *Neural networks for pattern recognition.* Oxford Univ. Press, Oxford, 1995.

[13] G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis.* John Wiley and Sons, New York, 1973,1992.

[14] P.F. Brown. *The Acoustic-Modeling Problem in Automatic Speech Recognition.* PhD thesis, Department of Computer Science, Carnegie Mellon University, 1987.

[15] U. Bub, J. Köhler, and B. Imperl. In-service adaptation of multilingual hidden-Markov-models. In *Proc. ICASSP '97*, pages 1451 – 1454, Munich, Germany, April 1997.

[16] D. Burshtein. Robust parametric modeling of durations in hidden Markov models. In *Proc. ICASSP '95*, pages 548 – 551, Detroit, MI, May 1995.

[17] P.-C. Chang and B.-H. Juang. Discriminative template training for dynamic programming speech recognition. In *Proc. ICASSP '92*, pages 493–496, San Francisco, CA, March 1992.

[18] J.-K. Chen and F. K. Soong. Discriminative training of high performance speech recognizer using N-best candidates. In *Proc. ICASSP '94*, pages I–625 – I–628, Adelaide, Australia, April 1994.

[19] J.-K. Chen and F.K. Soong. An N-best candidates-based discriminative training for speech recognition applications. *IEEE Transactions on Speech and Audio Processing*, 2(1):206–216, January 1994.

[20] J. T. Chien, H.-C. Wang, and C. H. Lee. Improved Bayesian learning of hidden Markov models for speaker adaptation. In *Proc. ICASSP '97*, pages 1027 – 1030, Munich, Germany, April 1997.

[21] J.T. Chien. In-line hierarchical transformation of hidden Markov models for speaker adaptation. In *Proc. ICSLP '98*, Sydney, Australia, November 1998.

[22] W. Chou. Maximum a posterior linear regression with elliptically symmetric matrix variate priors. In *Proc. Eurospeech '99*, pages 1–4, Budapest, Hungary, September 1999.

[23] W. Chou, B.-H. Juang, and C.-H. Lee. Segmental GPD training of HMM based speech recognizer. In *Proc. ICASSP '92*, pages 473–476, San Francisco, 1992. IEEE.

[24] W. Chou, C.-H. Lee, and B.-H. Juang. Minimum error rate training based on N-best string models. In *Proc. ICASSP '93*, pages I–652 – I–655, Minneapolis, MN, April 1993.

[25] C. Corredor-Ardoy, L. Lamel, M. Adda-Decker, and J-L. Gauvain. Multilingual phone recognition of spontaneous telephone speech. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[26] M. Creutz. Global Monte Carlo algorithms for many-fermion systems. *Physical Review D*, 38(4):1228–1238, Aug 1988.

[27] M.H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society series B*, 39:1–38, 1977.

[29] L. Deng. Integrated-multilingual speech recognition using universal phonological features in a functional speech production model. In *Proc. ICASSP '97*, pages 1007 – 1010, Munich, Germany, April 1997.

[30] V. Diakoloukas, V. Digalakis, L. Neumeyer, and J. Kaja. Development of dialect-specific speech recognizers using adaptation methods. In *Proc. ICASSP '97*, pages 1455 – 1458, Munich, Germany, April 1997.

[31] V.D. Diakoloukas and V.V. Digalakis. Maximum-likelihood stochachastic-transformation adaptation of hidden Markov models. *IEEE Transactions on Speech and Audio Processing*, 7(2):177–187, March 1999.

[32] V. Digalakis and L. Neumeyer. Speaker adaptation using combined transformation and Bayesian methods. In *Proc. ICASSP '95*, pages 680 – 683, Detroit, MI, May 1995.

[33] V.V. Digalakis, D. Rtischev, and L.G. Neumeyer. Speaker adaptation using constrained estimation of Gaussian mixtures. *IEEE Transactions on Speech and Audio Processing*, 3(5):357–366, September 1995.

[34] P.M. Djurić and J.H. Chun. Estimation of nonstationary hidden Markov models by MCMC sampling. In *Proc. ICASSP '99*, Phoenix, Arizona, USA, May 1999.

[35] J.A. Du Preez. Modelling durations in hidden Markov models with application to word spotting. In *Proc. IEEE South African symposium on Communications and Signal Processing*, pages 1–5, Fourways, South Africa, Aug. 1991.

[36] S. Duane. Stochastic quantization versus the microcanonical ensemble: getting the best of both worlds. *Nuclear Physics*, B257:652–662, 1985.

[37] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, Sep 1987.

[38] S. Duane and J.B. Kogut. Hybrid stochastic differential equations applied to quantum chromodynamics. *Physical Review Letters*, 55(25):2774–2777, Dec 1985.

[39] S. Duane and J.B. Kogut. The theory of hybrid stochastic algorithms. *Nuclear Physics*, B275:398–420, 1986.

[40] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis.* Wiley, New York, 1973.

[41] V. Fischer, Y. Gao, and E. Janke. Speaker-independent upfront dialect adaptation in a large vocabulary continuous speech recognizer. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[42] H. Franco and A. Serralheiro. Training HMMs using a minimum recognition approach. In *Proc. ICASSP '91*, pages 357–360, Toronto, Canada, May 1991.

[43] J-L. Gauvain and C.H. Lee. Bayesian learning for hidden Markov models with Gaussian mixture state observation densities. *Speech Communication*, pages 205–213, June 1992.

[44] J-L. Gauvain and C.H. Lee. Improved accoustic modeling with Bayesian learning. In *Proc. ICASSP '93*, pages II–481–II–484, Minneapolis, MN, April 1993.

[45] J-L. Gauvain and C.H. Lee. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291–298, April 1994.

[46] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1992.

[47] J. Glass, G. Flammia, D. Goodine, M. Phillips, J. Polifroni, S. Sakai, S. Seneff, and V. Zue. Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communication*, 17:1–18, August 1995.

[48] S.J. Godsill and C. Andrieu. Bayesian seperation and recovery of convolutively mixed autoregressive sources. In *Proc. ICASSP '99*, Phoenix, Arizona, USA, May 1999.

[49] Y. Gong and J.-P. Haton. Stochastic trajectory modeling for speech recognition. In *Proc. ICASSP '94*, pages I–57 – I–60, Adelaide, Australia, April 1994.

[50] R. Gupta, G.W. Kilcup, and S. R. Sharpe. Tuning the hybrid Monte Carlo algorithm. *Physical Review D*, 38(4):1278–1287, Aug 1988.

[51] S. K. Gupta, F. Soong, and R. Haimi-Cohen. High-accuracy connected digit recognition for mobile applications. In *Proc. ICASSP '96*, pages 57 – 60, Atlanta, GA, May 1996.

[52] Q. Huo and C. Chan. On-line Bayes adaptation of SCHMM parameters for speech recognition. In *Proc. ICASSP '95*, pages 708 – 711, Detroit, MI, May 1995.

[53] Q. Huo, C. Chan, and C.-H. Lee. Bayesian learning of the SCHMM parameters for speech recognition. In *Proc. ICASSP '94*, pages I–221 – I–224, Adelaide, Australia, April 1994.

[54] Q. Huo, H. Jiang, and C. H. Lee. A Bayesian predictive classification approach to robust speech recognition. In *Proc. ICASSP '97*, pages 1547 – 1550, Munich, Germany, April 1997.

[55] I. Illina and Y. Gong. Elimination of trajectory folding phenomenon: Trajectory mixture and mixture stochastic trajectory model. In *Proc. ICASSP '97*, pages 1395 – 1398, Munich, Germany, April 1997.

[56] C. Jankowski, A. Kalyanswamy, S. Basson, and J. Spitz. NTIMIT: A phonetically balanced, continuous speech, telephone bandwidth speech database. In *Proc. ICASSP '90*, pages 109–112, Alburquerque, NM, April 1990.

[57] H. Jiang, K. Hirose, and Q. Huo. Robust speech recognition based on Viterbi Bayesian predictive classification. In *Proc. ICASSP '97*, pages 1551 – 1554, Munich, Germany, April 1997.

[58] H. Jiang, K. Hirose, and Q. Huo. Improving Viterbi Bayesian predictive classification via sequential Bayesian learning in robust speech recognition. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[59] B.-H. Juang, W. Chou, and C.-H. Lee. Minimum classification error rate methods for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 5(3):257–265, May 1997.

[60] B.-H. Juang and S. Katagiri. Discriminative learning for minimum error classification. *IEEE Transactions on Signal Processing*, 40(12):3043–3054, December 1992.

[61] B.-H. Juang and L. R. Rabiner. The segmental K-means algorithm for estimating parameters of hidden Markov models. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-38(9), 1990.

[62] S. Kapadia, V. Valtchev, and S. J. Young. MMI training for continuous phoneme recognition on the TIMIT database. In *Proc. ICASSP '93*, Minneapolis, MN, April 1993.

[63] J. Köhler. Language adaptation of multilingual phone models for vocabulary independent speech recognition tasks. In *Proc. ICASSP '98*, pages 417 – 420, Seattle, USA, May 1998.

[64] T. Komori and S. Katagiri. Application of a generalized probabilistic descent method to dynamic time warping-based speech recognition. In *Proc. ICASSP '92*, pages 497–500, San Francisco, CA, March 1992.

[65] O.W. Kwon and C.K. Un. Performance of HMM-based speech recognizers with discriminative state-weights. *Speech Communication*, 19:197–205, 1996.

[66] C-H. Lee and J-L. Gauvain. Speaker adaptation based on MAP estimation of HMM parameters. In *Proc. ICASSP '93*, pages II–558–II–561, Minneapolis, MN, April 1993.

[67] C-H. Lee, C-H Lin, and B-H Juang. A study on speaker adaptation of the parameters of continuous density hidden Markov models. *IEEE Trans. Signal Processing*, 39(4):806–814, April 1991.

[68] K. F. Lee and H. W. Hon. Speaker-independent phone recognition using hidden Markov models. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-37(11):1641, 1989.

[69] C. J. Leggetter and P. C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*, 9(2):171–185, April 1995.

[70] R. G. Leonard. A database for speaker-independant digit recognition. In *Proc. ICASSP '84*, pages 42.11.1–42.11.4, 1984.

[71] L. A. Liporace. Maximum likelihood estimation for multivariate observations of Markov sources. *IEEE Transactions on Information Theory*, 28(5):729–734, Spetember 1982.

[72] David J.C. Mackay. *Bayesian Methods for Adaptive Models.* PhD thesis, California Institute of Technology, Pasadena, California, December 1991.

[73] J. S. Maritz and T. Lwin. *Empirical Bayes Methods.* Chapman and Hall, London, 1989.

[74] T. Matsui and S. Furui. A study of speaker adaptation based on minimum classification error training. In *Proc. Eurospeech '95*, pages 81–84, Madrid, Spain, September 1995.

[75] E. McDermott. *Discriminative Training for Speech Recognition.* PhD thesis, Waseda University, Japan, March 1997.

[76] E. McDermott and S. Katagiri. Prototype-based discriminative training for various speech units. In *Proc. ICASSP '92*, San Francisco, CA, March 1992.

[77] E. McDermott and S. Katagiri. Prototype-based MCE/GPD training for word spotting and connected word recognition. In *Proc. ICASSP '93*, pages II–291 – II–294, Minneapolis, MN, April 1993.

[78] E. McDermott and S. Katagiri. String-level MCE for continuous phoneme recognition. In *Proc. Eurospeech '97*, pages 123–126, Rhodes, Greece, September 1997.

[79] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, and A.H. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, Jun 1953.

[80] P. J. Moreno and R. M. Stern. Sources of degradation of speech recognition in telephone environments. In *Proc. ICASSP '94*, pages I–109 – I–112, Adelaide, Australia, April 1994.

[81] A. Nadas, D. Nahamoo, and M. A. Picheny. On a model-robust training method for speech recognition. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-36(9):1432, 1988.

[82] R.M. Neal. *Bayesian Learning for Neural Networks*. Springer, New York, 1996.

[83] C. Nieuwoudt. *Cross-language acoustic adaptation for automatic speech recognition*. PhD thesis, University of Pretoria, South Africa, April 2000.

[84] C. Nieuwoudt and E.C. Botha. Adaptation of acoustic models for multilingual recognition. In *Proc. Eurospeech '99*, pages 907–910, Budapest, Hungary, September 1999.

[85] C. Nieuwoudt and E.C. Botha. Cross-language use of accoustic information for automatic speech recognition. In *Proc. ICSLP 2000*, volume 3, pages 722–725, Beijing, China, October 2000.

[86] N. Nilsson. *Problem-Solving Methods in Artificial Intelligence*. McGraw Hill, New York, New York, 1971.

[87] Y. Normandin. Optimal splitting of HMM gaussian mixture components with MMIE training. In *Proc. ICASSP '95*, pages 449 – 452, Detroit, MI, May 1995.

[88] Y. Normandin and S.D. Morgera. An improved MMIE training algorithm for speaker-independent, small vocabulary, continuous speech recognition. In *Proc. ICASSP '91*, pages 537–540, Toronto, Canada, May 1991.

[89] D.W. Purnell and E.C. Botha. Improved performance and generalization of minimum classification error training for continuous speech recognition. In *Proc. ICSLP 2000*, volume 4, pages 165–168, Beijing, China, October 2000.

[90] L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition.* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[91] L. R. Rabiner, J. G. Wilpon, and F. K. Soong. High performance connected digit recognition using hidden Markov models. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-37(8):1214, 1989.

[92] L.R. Rabiner. An introduction to hidden Markov models. *IEEE Signal Processing Magazine*, 5:4–16, January 1988.

[93] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77:257–286, Feb 1989.

[94] L.R. Rabiner and B.H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, pages 4–16, January 1986.

[95] C. Rathinavelu. Minimum classification error linear regression (MCELR) for speaker adaptation using HMM with trend functions. In *Proc. Eurospeech '97*, pages 2343–2346, Rhodes, Greece, September 1997.

[96] C. Rathinavelu and L. Deng. HMM-based speech recognition using state-dependent linear transforms on mel-warped DFT features. In *Proc. ICASSP '96*, pages 9 - 12, Atlanta, GA, May 1996.

[97] H. Robbins. The empirical Bayes approach to statistical decision problems. *Annals of Mathematical Statistics*, 35:1–20, 1964.

[98] C.P. Robert, A. Doucet, and S.J. Godsill. Marginal MAP estimation using Markov chain Monte Carlo. In *Proc. ICASSP '99*, Phoenix, Arizona, USA, May 1999.

[99] S. Russel and P. Norvig. *Artificial Intelligence, A Modern Approach.* Prentice Hall, Upper Saddle River, New Jersey, 1995.

[100] A. Sankar. Robust HMM estimation with Gaussian merging-splitting and tied-transform HMMs. In *Proc. ICSLP '98*, volume 6, pages 2499–2502, Sydney, Australia, November 1998.

[101] R. Schlüter and W. Machery. Comparison of discriminative training criteria. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[102] T. Schultz and A. Waibel. Fast bootstrapping of LVCSR systems with multilingual phoneme sets. In *Proc. Eurospeech '97*, pages 371–374, Rhodes, Greece, September 1997.

[103] R. Schwartz and S. Austin. A comparision of several approximate algorithms for finding multiple (NBEST) sentence hypotheses. In *Proc. ICASSP '91*, pages 701–704, Toronto, Canada, May 1991.

[104] H. Shimodaira, J. Rokui, and M. Nakai. Improving the generalization performance of the MCE/GPD learning. In *Proc. ICSLP '98*, Sydney, Australia, November 1998.

[105] O. Siohan, A.E. Rosenberg, and S. Parthasarathy. Speaker identification using minimum classification error training. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[106] F. K. Soong and E.-F. Huang. A tree-trellis based fast search for finding the N best sentence hypotheses in continuous speech recognition. In *Proc. ICASSP '91*, pages 705–708, Toronto, Canada, May 1991.

[107] J. Su, H. Li, J.-P. Haton, and K.-T. Ng. Speaker time-drifting adaptation using trajectory mixture hidden Markov models. In *Proc. ICASSP '96*, pages 709 – 712, Atlanta, GA, May 1996.

[108] R. A. Sukkar. Rejection for connected digit recognition based on GPD segmental discrimination. In *Proc. ICASSP '94*, pages I–393 – I–396, Adelaide, Australia, April 1994.

[109] R.A. Sukkar and J.G. Wilpon. A two pass classifier for utterance rejection in keyword spotting. In *Proc. ICASSP '93*, pages II–451 – II–454, Minneapolis, MN, April 1993.

[110] E. Thelen, X. Aubert, and P. Beyerlein. Speaker adaptation in the Philips system for large vocabulary continuous speech recognition. In *Proc. ICASSP '97*, pages 1035 – 1038, Munich, Germany, April 1997.

[111] D. Toussaint. Introduction to algorithms for Monte Carlo simulations and their application to QCD. *Computer Physics Communications*, 56:69–92, 1989.

[112] U. Uebler, M. Schussler, and H. Niemann. Bilingual and dialectal adaptation and retraining. In *Proc. ICSLP '98*, volume 5, pages 1815–1818, Sydney, Australia, November 1998.

[113] V. Valtchev, J.J. Odell, P.C. Woodland, and S.J. Young. MMIE training of large vocabulary recognition systems. *Speech Communication*, 22:303–314, 1997.

[114] J. Vermaak and M. Niranjan. Markov chain Monte Carlo methods for speech enhancement. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[115] T. Waardenburg, J.A. du Preez, and M.W. Coetzer. The automatic recognition of stop consonants using hidden Markov models. In *Proc. ICASSP '92*, San Francisco, CA, March 1992.

[116] B. Wheatley, K. Kondo, W. Anderson, , and Y. Muthusamy. An evaluation of cross-language adaptation for rapid HMM development in a new language. In *Proc. ICASSP '94*, pages I–237 – I–240, Adelaide, Australia, April 1994.

[117] P. Woodland, J. Odell, V. Valtchev, and S. Young. Large vocabulary continuous speech recognition using HTK. In *Proc. ICASSP '94*, pages II–125 – II–128, Adelaide, Australia, April 1994.

[118] P.C. Woodland, T. Hain, S.E. Johnson, T.R. Niesler, A. Tuerk, and S.J. Young. Experiments in broadcast news transcription. In *Proc. ICASSP '98*, Seattle, USA, May 1998.

[119] D. Yuk and J. Flanagan. Telephone speech recognition using neural networks and Markov models. In *Proc. ICASSP '98*, pages 1872–1875, Seattle, USA, May 1998.

[120] G. Zavaliagkos, R. Schwartz, and J. Makhoul. Batch, incremental and instantaneous adaptation techniques for speech recognition. In *Proc. ICASSP '95*, pages 676 – 679, Detroit, MI, May 1995.

Appendix A

Probability distributions

A.1 The normal distribution

The p.d.f. of a k-dimensional multivariate normal distribution is specified as follows [27]:

$$g(x|\mu,\Sigma) = (2\pi)^{-D/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}. \tag{A.1}$$

A normal distribution with mean μ and covariance matrix Σ is usually written as $\mathcal{N}(\mu,\Sigma)$. If we assume that the individual elements of the random variable x are independent (diagonal covariance matrix), we can write Equation A.1 as follows

$$g(x|\mu,\Sigma) = (2\pi)^{-D/2}\left(\prod_{d=1}^{D}\sigma_d\right)^{-1} e^{-\frac{1}{2}\sum_{d=1}^{D}\sigma_d^{-2}(x_d-\mu_d)^2} \tag{A.2}$$

and

$$log(g(x|\mu,\Sigma)) = -\frac{D}{2}log(2\pi) - \frac{1}{2}log\left(\prod_{d=1}^{D}\sigma_d^2\right) - \frac{1}{2}\sum_{d=1}^{D}\sigma_d^{-2}(x_d-\mu_d)^2 \qquad (A.3)$$

A.2 The Wishart distribution

Given a random sample of k-dimensional random vectors $(X_1, X_2, \ldots, X_n)$ from a multivariate normal distribution with zero mean and covariance matrix Σ, the random variable V has a Wishart distribution [27] with n degrees of freedom and parametric matrix Σ when,

$$V = \sum_{i=1}^{n} X_i X_i^T. \qquad (A.4)$$

For any $k \times k$, positive definite, symmetric matrix v, we have

$$g(v|n,\Sigma) = c|\Sigma|^{-n/2}|v|^{(n-D-1)/2}e^{-\frac{1}{2}tr(\Sigma^{-1}v)}. \qquad (A.5)$$

Here, $tr(\Sigma^{-1}v)$ is the trace of the matrix $\Sigma^{-1}v$. The value c is a normalizing constant which ensures that the integral of $g(v|n,\Sigma)$ is equal to one.

If Σ and v are assumed to be diagonal, then Equation A.5 can be rewritten as follows:

$$g(v|n,\Sigma) = c\left(\prod_{d=1}^{D}\tau_d\right)^{n/2}\left(\prod_{d=1}^{D}v_d\right)^{(n-D-1)/2} e^{-\frac{1}{2}\sum_{d=1}^{D}v_d\tau_d} \qquad (A.6)$$

where τ is the precision matrix for the Wishart distribution ($\tau = \Sigma^{-1}$) and

$$log(g(v|n,\Sigma)) = log(c) + \frac{n}{2}\cdot log\left(\prod_{d=1}^{D}\tau_d\right) + \frac{n-D-1}{2}log\left(\prod_{d=1}^{D}v_d\right) - \frac{1}{2}\sum_{d=1}^{D}v_d\tau_d. \quad (A.7)$$

A.3 Dirichlet distribution

Given the random vector $X = (X_1, X_2, \ldots, X_k)^T$ with the following properties: For a point $x = (x_1, x_2, \ldots, x_k)^T$ in $\Re^k$, $x_i > 0; i = 1, \ldots, k$ and $\sum_{i=1}^{k} x_i = 1$, then the random vector X has a Dirichlet distribution [27]:

$$g(x|\alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_k)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_k^{\alpha_k - 1}, \tag{A.8}$$

where $\Gamma(\alpha)$ is the gamma function and α is the parametric vector of the distribution and

$$\begin{aligned} log(g(x|\alpha)) =& log(\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)) - log(\Gamma(\alpha_1)) - \cdots - log(\Gamma(\alpha_k)) \\ &+ (\alpha_1 - 1)log(x_1) + \cdots + (\alpha_k - 1)log(x_k). \end{aligned} \tag{A.9}$$

A.4 The gamma distribution

A random variable X has a gamma distribution [27] with parameters α and β ($\alpha > 0, \beta > 0$) if X has an absolutely continuous distribution whose p.d.f. is

$$g(x|\alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.10}$$

where $\Gamma(\alpha)$ is the gamma function, which is defined as

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \tag{A.11}$$

If a random variable X has a gamma distribution as given in (A.10), then

$$E(X) = \frac{\alpha}{\beta} \tag{A.12}$$

$$Var(X) = \frac{\alpha}{\beta^2}. \tag{A.13}$$

If the mean μ and variance σ^2 of a gamma distribution are known, the distribution parameters α and β can easily be obtained as follows:

$$\alpha = \frac{\mu^2}{\sigma^2} \tag{A.14}$$

$$\beta = \frac{\mu}{\sigma^2}. \tag{A.15}$$

In this thesis, a gamma distribution with parameters α and β has been referred to as $\mathcal{G}(\alpha, \beta)$.

A.5 Conjugate families of distributions

If the prior distribution of θ belongs to a conjugate family of distributions [27], then for any sample size n and any values of the observations in the sample, the posterior distribution of θ must also belong to the same family. A family of distributions with this property is said to be closed under sampling.