Universidad Nacional de San Martín

ICAS

Técnicas estadísticas para el análisis de datos astrofísicos

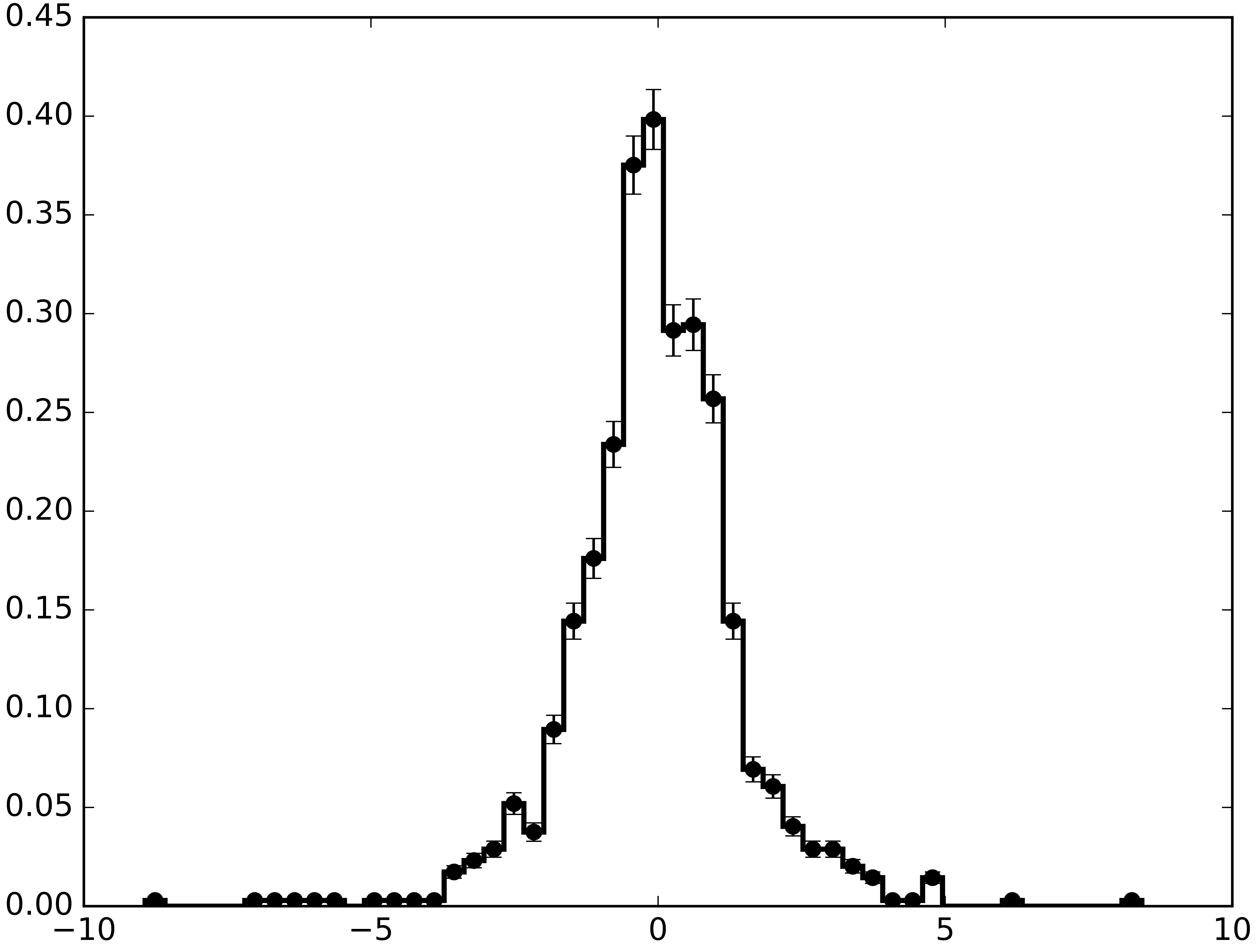
Rodrigo F. Díaz

International Center for Advanced Studies (ICAS)
University of San Martín (UNSAM), Argentina

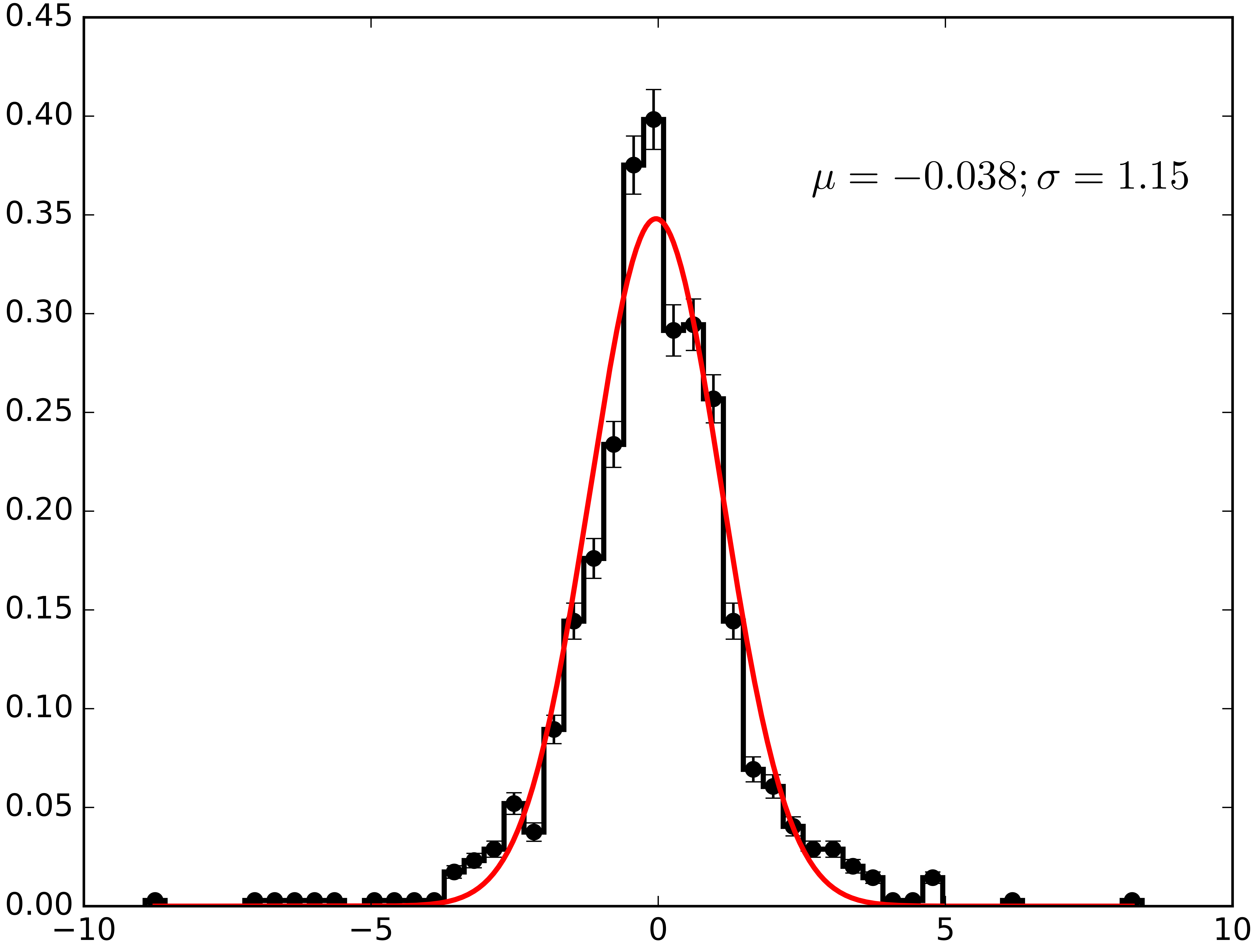
rdiaz@unsam.edu.ar

Primera Escuela Latinoamericana de Relatividad y Asrofísica
Diciembre de 2021

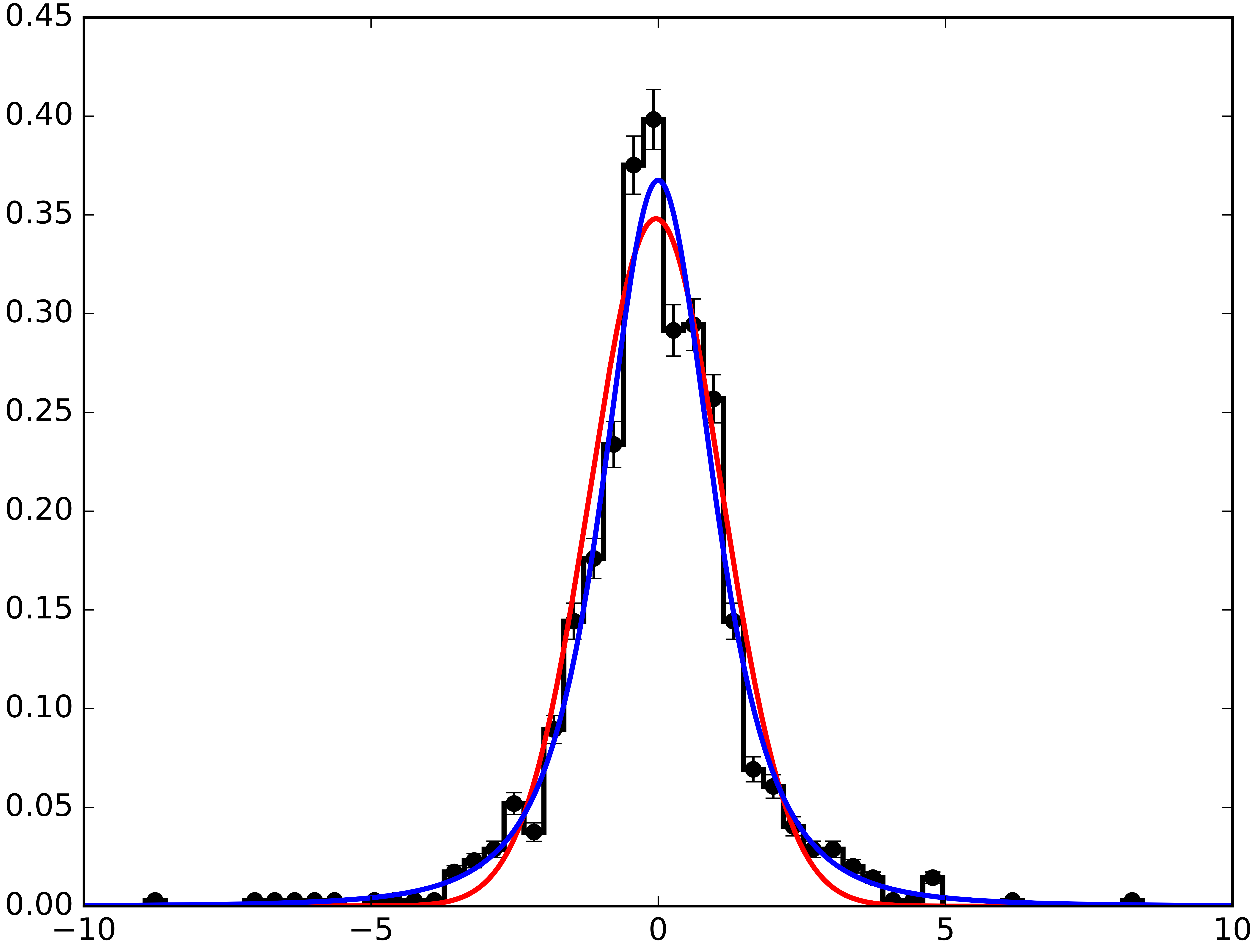
0.0766, 2.3952, 1.0476, -1.2858, -0.2306, 1.1802, -3.3299, -0.4132, -0.3664, -1.5994, -0.3148, 4.3496, -2.5871, 1.3305, 0.9727, 0.6839, 0.0276, 0.2045, 1.2147, 2.5365, 1.1441, 3.2629, 0.1362, -0.4081, -0.6274, 0.6130, -0.3459,
0.4389, -0.9244, -2.0190, -1.2441, 1.4978, 0.2328, -0.2747, -0.5876, 0.9303, 0.4465, -0.4086, -0.6834, -1.6257, -3.4097, -0.2872, 0.3217, -0.8903, 0.5751, -2.5417, 0.7618, -0.1990, 0.4749, -0.6886, 0.4808, -0.6274, -0.1490,
1.8698, 0.5414, 0.2706, 0.8269, -0.4423, 0.6014, -0.0533, -0.5480, 0.9465, -0.0274, -0.7399, 0.1950, -3.0122, -0.2117, 1.0063, -1.2950, -0.8230, -0.0239, 0.4498, -0.8517, 3.6613, 1.3914, 1.5189, -2.1787, 0.3822, 0.1255,
0.3432, 1.0617, -0.3204, 1.3348, 0.9355, 0.6301, 3.4833, -0.8114, 0.1575, 3.1708, 0.6257, 0.7065, -1.1157, -2.3580, -4.6061, 0.8672, 0.9294, 2.9604, 0.1266, -1.9851, -0.8187, -0.1706, 2.7169, 2.7969, 2.1371, 3.0910, -0.8833,
0.2133, -8.9333, -1.2552, -0.8953, -0.0167, 0.5444, 0.4873, -1.1230, -2.2517, -1.1703, 2.3537, 0.7819, 0.6830, 0.2425, -0.5108, 0.0541, 0.3107, 0.0824, -0.0954, -1.9473, -0.4189, 0.5788, -1.7985, 1.1069, -1.0192, -0.3612,
-0.1431, -0.3686, -0.3657, -1.3466, 1.0715, -0.0615, 0.7700, 0.9807, 0.6837, -0.4829, -0.4847, -0.0408, -2.9214, -0.7188, 0.4714, 0.2608, 0.1678, -0.7726, 2.4223, -1.6223, 1.1990, 0.0309, -3.1811, -0.0955, -1.4879, 0.8964,
-0.1063, -0.2639, 1.4454, 0.1860, -0.3108, 1.0299, 0.4298, -0.4844, -0.1514, -0.0100, -1.2320, 0.9244, 0.3842, -2.0463, -3.5029, 2.2090, -0.0722, -0.3270, -1.6422, -0.6720, 0.1335, 0.7404, -0.0754, -0.9890, 1.7827, 4.0538,
-0.0620, -0.5512, 0.8910, 0.6116, -0.9693, 0.2661, -0.4064, -2.1055, 0.4639, 0.3008, -0.1206, 0.1843, 3.2056, -3.3907, 0.8447, -0.6212, 0.2249, -1.9509, 0.5351, -1.2619, 1.8796, -0.7371, -0.6632, 0.9288, -2.0942, 1.1829,
-1.2170, -0.3879, 0.2870, -0.3780, 0.4796, 4.6440, -3.2195, -0.7724, -1.3707, -0.4496, 0.2207, -0.3735, 4.8009, 0.2859, 0.3271, -0.5053, -3.2197, 0.0770, -3.0705, -2.7360, 0.0243, 1.3920, -1.3508, 1.5471, 0.0783, -2.5703,
1.4966, -1.5306, -2.7655, -1.0365, -0.5016, 0.1065, 0.5727, 0.8461, -0.1113, -0.2310, 0.9931, -0.3884, 0.7402, -1.5023, -0.6913, -1.0615, -1.9504, 1.1166, 1.6918, -0.1562, -1.4498, 0.0006, 1.3752, -0.3413, -0.7429, 0.3983,
0.4719, -1.6959, -0.2056, 0.2178, 0.5141, -0.5745, 0.5449, 0.2420, -0.1296, -1.9691, -2.4805, 1.2441, 1.4198, -0.9786, 1.2632, 1.9183, -0.4359, -0.1306, 1.4974, 3.4456, 0.1820, 0.2420, -0.1202, -1.5408, -1.5500, -0.3047,
0.0402, 1.4828, -0.4459, 3.0487, 0.6775, -1.6424, 0.4681, -1.3879, 0.0298, 0.1707, -0.5672, -0.2238, 0.1543, 0.4965, -0.8725, -0.1862, 0.0536, 1.9399, 0.8416, 0.0080, 2.3084, 0.8265, 0.4797, 1.3048, -0.6953, -0.9466, 0.2398,
0.2648, 1.0545, -0.2481, -0.7224, 0.7350, 0.0967, 0.5514, 0.8454, 0.7592, 0.7669, -1.4729, -0.9691, 0.5400, 1.7118, 0.2743, 0.4738, -1.4815, 4.9319, -0.3526, 2.2348, -0.0072, -1.8503, -1.2781, -0.4065, 1.1592, -1.4303,
-1.1181, -1.1649, 1.0389, 0.9183, 2.8120, -1.0726, 2.3604, 0.7836, 0.8147, -0.8836, 1.1037, 0.2468, -5.7165, 0.2066, 0.5739, -0.5308, -0.5953, 0.2851, -1.3518, -1.2859, 0.7298, -1.5753, -1.3553, 3.4118, 1.8400, -1.0751,
0.7278, -0.2014, -0.1527, -0.2745, -0.4837, -0.6613, -3.5672, 0.2222, 0.9499, 2.1372, 1.4601, -1.7270, -0.2280, 1.6503, -0.0368, -5.9512, 1.9027, 1.5553, 1.2709, 1.4792, -0.2827, -0.1447, 0.9484, -0.6560, -0.1931, -0.4891,
1.2324, -0.2650, 2.6470, -1.5580, 1.1036, -2.4841, -0.2640, -0.2440, 0.6009, -0.1926, 1.4688, -0.6251, 3.3118, -0.4526, 1.0283, -0.3820, -3.0154, -1.1445, 0.8195, -0.3249, -0.3703, 3.5886, -0.7405, -1.2940, 1.2646, -0.3361,
-1.7244, 0.2509, 2.9173, -0.1308, 0.9136, 0.0706, -0.9477, -2.4055, 0.7462, 0.9538, -0.8179, -0.4440, -1.1852, -0.7789, -0.5011, -1.8987, -0.3188, 2.0584, 0.6024, -3.1321, -0.6272, -0.5745, 0.4365, 0.9436, -0.6201, 1.3551,
-0.4047, 1.5559, 1.0733, -0.3982, -1.4490, -0.0592, 2.8637, 0.9025, -0.7550, 1.5398, -0.5594, 0.4108, -0.5978, 0.0613, -0.9095, -0.4517, -0.9100, -0.3021, -0.3001, -0.2229, 0.9803, 1.8816, 2.9271, 3.3045, -1.2201, -0.9750,
0.0013, -0.9089, -0.2484, 0.0234, -1.1106, -0.8048, 1.0323, -0.5966, 6.1361, -0.1659, -0.0031, 0.0573, -2.0689, -1.6038, 0.6123, 0.4766, -0.1283, -0.3109, -6.5703, 3.0460, -0.4466, -0.3737, -0.1055, -0.1960, -1.8886, 0.9178,
-1.5232, 0.0379, -1.4492, -1.1803, 1.0930, -0.0066, 0.2857, -0.0220, -1.0741, 0.5158, 1.5005, 0.7451, 0.4993, -0.0173, -0.5180, -0.3155, -0.9085, -0.0370, 2.1473, -0.6131, -1.8105, 0.3878, 1.2429, -0.3604, -0.3411, 2.8003,
-0.5535, -2.9557, 3.8684, -2.5195, -0.9888, 1.3046, 2.0976, 0.1146, 1.1547, 0.7065, -0.9618, -0.9320, 0.6342, -0.4095, -0.4565, -1.5873, -0.2245, 2.5148, -6.8740, -0.5089, -0.1434, -0.4689, 1.6055, -0.2133, -0.4027, 3.5856,
-0.0116, -1.7678, 0.0375, 0.0212, 0.0312, -0.8189, 1.3116, 0.9817, 1.8707, 0.0307, -0.6598, -0.5693, 0.2901, -0.1544, -0.7304, -0.8341, 1.1035, 0.4524, -0.5887, 1.7743, 0.8751, -0.2558, 0.4451, 0.4259, 0.0815, -0.5457,
-0.3495, -0.8443, -0.3872, -2.6320, 0.8772, -1.7566 -0.4258, 0.9629, 0.8271, -0.0761, -0.1620, -0.2599, 0.6481, 1.2517, 0.2244, 0.0331, -1.0516, 2.8017, 1.8559, 0.9382, 0.8312, -0.1303, 0.9221, 0.6060, -0.5698, -0.3927,
-1.3152, 1.1827, 1.6165, -0.6727, 0.8533, -0.1789, -1.2074, 0.8020, -0.1568, 0.7057, 0.5656, -0.7147, 1.1954, 0.5416, -0.2113, 1.0522, 1.9738, 0.1474, 0.6597, 0.3949, -1.2146, 0.2744, 0.6703, -0.6913, -0.9657, 0.8498, 0.9167,
-0.5060, 0.7082, 0.3312, -0.3092, -0.3044, -1.5772, -0.9272, -0.8858, 1.3993, -0.3382, -2.8431, 0.9452, -3.0444, -2.0309, 0.1416, 0.1474, -0.0208, -2.9255, 0.3000, -3.8650, -0.0282, 1.1727, 0.5815, -1.1257, 8.4252, -2.5299,
0.0382, -0.6373, 0.2693, -0.3993, 0.6195, 1.2367, 0.8516, -0.5885, 0.6170, 1.3144, 4.7261, -1.7593, -0.0083, 0.4488, -0.8542, -0.8866, 0.9777, -0.7837, -1.1594, -4.0903, 0.3139, 1.3620, -0.5746, 1.2583, -0.4483, -0.0271,
-0.5161, -1.8327, 3.2739, -0.2937, -0.1833, -1.2682, 0.5631, 2.1498, -0.5458, -0.7951, -0.9052, -0.9792, -1.6048, -1.3072, -3.5114, -0.1072, -0.5162, -1.2271, -1.2452, -0.3166, -4.9986, 1.4746, -0.4988, -0.0378, 0.0032,
0.2864, -0.4586, -0.1442, 0.4167, 0.1680, 0.4311, 1.2459, 0.4811, 2.0354, -0.1447, 0.0320, 0.0170, 1.1043, 0.0977, -1.0106, 2.1566, -0.3881, 0.8465, 0.7347, -0.6717, -0.7952, 0.0710, 3.6445, -0.0550, -0.3550, 0.6799, -1.1895,
-0.0828, 1.3692, -2.1940, -0.5108, 0.4514, -1.5967, -1.0766, -0.0951, -2.6083, -0.2892, -1.3840, 0.3676, 1.1254, -0.5696, -0.4879, 0.0799, -1.1759, 0.9598, 1.9274, 0.7494, 0.1644, -1.6152, -2.5340, 0.3410, 1.0850, 0.5481,
-0.1523, -2.2027, -2.5697, -1.0047, 0.5154, 0.8907, 0.2663, 0.7922, 1.0089, -1.4280, 2.2191, 0.8956, -0.4291, 0.4167, -0.9900, -0.0448, -2.0657, 0.0179, 0.6709, -1.4507, 0.0385, -0.8988, 0.9545, 0.9640, -1.7518, -1.0884,
-0.4236, 1.6847, -1.9293, 0.6477, 0.4632, 4.8035, -0.4204, -0.7591, 0.4864, 0.9606, -1.7324, 0.3425, -0.8570, -0.7900, 1.0013, -1.4189, -1.6568, -0.2089, 0.0437, 0.3574, 0.3368, -1.0191, 1.3581, -1.1470, -1.7700, -0.2553,
0.2416, 1.4234, 0.0679, -2.3670, -0.4121, 0.5269, -0.2469, 0.0158, -0.1713, 1.1389, 1.1288, -0.4167, -0.7257, -0.5442, 0.0650, 0.6510, -1.5929, 1.0691, 0.3750, -2.2971, 2.9754, 2.9149, -1.3911, 2.2526, 0.1382, -1.7603,
-1.0631, 0.5170, -0.1009, 1.1263, -0.0092, -0.0098, -2.7842, 0.6913, -0.9865, 0.2700, 0.9492, -0.0675, 1.0261, 0.0933, -0.9754, -0.9568, -0.2461, -1.5360, 0.4426, 0.2175, 0.7143, -0.5052, -0.8001, 0.5656, 0.7063, 1.4770,
-1.3006, -1.5714, -0.2693, 2.2670, 0.1558, 0.1991, -1.6480, 0.5153, -2.4253, 1.5756, 0.5443, 2.5773, -1.9362, -0.2781, -2.0731, 0.1090, -0.5054, -0.4368, -0.7355, -0.1636, -1.6857, -1.5232, 1.8666, 0.7096, 1.6585, -0.2773,
1.0332, 1.6389, -1.6783, 0.4412, 0.2219, 0.3065, 0.2976, -0.5809, 2.1519, 0.0975, 1.0190, -0.0502, -2.4786, -1.7093, 1.0806, -3.3706, 1.6150, -0.2482, -0.3966, -0.6100, 0.5156, -0.5557, 0.7086, 1.2471, 0.3134, -1.7028, 0.2730,
1.3062, -2.3565, -0.8197, 1.0612, 1.2029, 0.9465, 0.3207, -0.2665, -0.7594, -1.0070, 2.2674, 0.6885, -1.3172, -0.5294, 0.3230, -0.8870, -1.6809, 0.3694, -0.8466, -0.7692, 0.8474, 1.0270, 0.5491, -1.0072, 2.5180, -0.6298,
-1.5580, 0.7622, 1.3980, 0.3573, -0.2377, -0.1133, -3.4173, -1.2886, -0.6681, -1.9261, 0.5664, 0.0513, -1.0560, -0.0775, -0.0587, -1.5512, 2.4590, 0.4063, 1.4058, 0.0397, 0.1408, -0.3774, 0.3097, -0.7795, -1.3605, -1.3067,
-0.7253, -6.3967, 1.2869, -1.3681, 0.3538, -0.0138, 0.9987, -0.0393, -0.0788, -2.8061, -0.6761, -0.2404, -0.6004, 1.5730, 2.5828, 0.2641, 0.1437, -0.0809, -1.6422, 0.5712, 0.8864, 0.5777, 0.6425, -0.0133, -2.3401, 1.6605,
1.4231, 0.8345, -0.7660, -1.1168, -0.5035, 0.0766, 2.3952, 1.0476, -1.2858, -0.2306, 1.1802, -3.3299, -0.4132, -0.3664, -1.5994, -0.3148, 4.3496, -2.5871, 1.3305, 0.9727, 0.6839, 0.0276, 0.2045, 1.2147, 2.5365, 1.1441, 3.2629,
0.1362, -0.4081, -0.6274, 0.6130, -0.3459, 0.4389, -0.9244, -2.0190, -1.2441, 1.4978, 0.2328, -0.2747, -0.5876, 0.9303, 0.4465, -0.4086, -0.6834, -1.6257, -3.4097, -0.2872, 0.3217, -0.8903, 0.5751, -2.5417, 0.7618, -0.1990,
0.4749, -0.6886, 0.4808, -0.6274, -0.1490, 1.8698, 0.5414, 0.2706, 0.8269, -0.4423, 0.6014, -0.0533, -0.5480, 0.9465, -0.0274, -0.7399, 0.1950, -3.0122, -0.2117, 1.0063, -1.2950, -0.8230, -0.0239, 0.4498, -0.8517, 3.6613,
1.3914, 1.5189, -2.1787, 0.3822, 0.1255, 0.3432, 1.0617, -0.3204, 1.3348, 0.9355, 0.6301, 3.4833, -0.8114, 0.1575, 3.1708, 0.6257, 0.7065, -1.1157, -2.3580, -4.6061, 0.8672, 0.9294, 2.9604, 0.1266, -1.9851, -0.8187, -0.1706,
2.7169, 2.7969, 2.1371, 3.0910, -0.8833, 0.2133, -8.9333, -1.2552, -0.8953, -0.0167, 0.5444, 0.4873, -1.1230, -2.2517, -1.1703, 2.3537, 0.7819, 0.6830, 0.2425, -0.5108, 0.0541, 0.3107, 0.0824, -0.0954, -1.9473, -0.4189,
0.5788, -1.7985, 1.1069, -1.0192, -0.3612, -0.1431, -0.3686, -0.3657, -1.3466, 1.0715, -0.0615, 0.7700, 0.9807, 0.6837, -0.4829, -0.4847, -0.0408, -2.9214, -0.7188, 0.4714, 0.2608, 0.1678, -0.7726, 2.4223, -1.6223, 1.1990,
0.0309, -3.1811, -0.0955, -1.4879, 0.8964, -0.1063, -0.2639, 1.4454, 0.1860, -0.3108, 1.0299, 0.4298, -0.4844, -0.1514, -0.0100, -1.2320, 0.9244, 0.3842, -2.0463, -3.5029, 2.2090, -0.0722, -0.3270, -1.6422, -0.6720, 0.1335,
0.7404, -0.0754, -0.9890, 1.7827, 4.0538, -0.0620, -0.5512, 0.8910, 0.6116, -0.9693, 0.2661, -0.4064, -2.1055, 0.4639, 0.3008, -0.1206, 0.1843, 3.2056, -3.3907, 0.8447, -0.6212, 0.2249, -1.9509, 0.5351, -1.2619, 1.8796,
-0.7371, -0.6632, 0.9288, -2.0942, 1.1829, -1.2170, -0.3879, 0.2870, -0.3780, 0.4796, 4.6440, -3.2195, -0.7724, -1.3707, -0.4496, 0.2207, -0.3735, 4.8009, 0.2859, 0.3271, -0.5053, -3.2197, 0.0770, -3.0705, -2.7360, 0.0243,
1.3920, -1.3508, 1.5471, 0.0783, -2.5703, 1.4966, -1.5306, -2.7655, -1.0365, -0.5016, 0.1065, 0.5727, 0.8461, -0.1113, -0.2310, 0.9931, -0.3884, 0.7402, -1.5023, -0.6913, -1.0615, -1.9504, 1.1166, 1.6918, -0.1562, -1.4498,
0.0006, 1.3752, -0.3413, -0.7429, 0.3983, 0.4719, -1.6959, -0.2056, 0.2178, 0.5141, -0.5745, 0.5449, 0.2420, -0.1296, -1.9691, -2.4805, 1.2441, 1.4198, -0.9786, 1.2632, 1.9183, -0.4359, -0.1306, 1.4974, 3.4456, 0.1820, 0.2420,
-0.1202, -1.5408, -1.5500, -0.3047, 0.0402, 1.4828, -0.4459, 3.0487, 0.6775, -1.6424, 0.4681, -1.3879, 0.0298, 0.1707, -0.5672, -0.2238, 0.1543, 0.4965, -0.8725, -0.1862, 0.0536, 1.9399, 0.8416, 0.0080, 2.3084, 0.8265, 0.4797,
1.3048, -0.6953, -0.9466, 0.2398, 0.2648, 1.0545, -0.2481, -0.7224, 0.7350, 0.0967, 0.5514, 0.8454, 0.7592, 0.7669, -1.4729, -0.9691, 0.5400, 1.7118, 0.2743, 0.4738, -1.4815, 4.9319, -0.3526, 2.2348, -0.0072, -1.8503, -1.2781,
-0.4065, 1.1592, -1.4303, -1.1181, -1.1649, 1.0389, 0.9183, 2.8120, -1.0726, 2.3604, 0.7836, 0.8147, -0.8836, 1.1037, 0.2468, -5.7165, 0.2066, 0.5739, -0.5308, -0.5953, 0.2851, -1.3518, -1.2859, 0.7298, -1.5753, -1.3553,
3.4118, 1.8400, -1.0751, 0.7278, -0.2014, -0.1527, -0.2745, -0.4837, -0.6613, -3.5672, 0.2222, 0.9499, 2.1372, 1.4601, -1.7270, -0.2280, 1.6503, -0.0368, -5.9512, 1.9027, 1.5553, 1.2709, 1.4792, -0.2827, -0.1447, 0.9484,
-0.6560, -0.1931, -0.4891, 1.2324, -0.2650, 2.6470, -1.5580, 1.1036, -2.4841, -0.2640, -0.2440, 0.6009, -0.1926, 1.4688, -0.6251, 3.3118, -0.4526, 1.0283, -0.3820, -3.0154, -1.1445, 0.8195, -0.3249, -0.3703, 3.5886, -0.7405,
-1.2940, 1.2646, -0.3361, -1.7244, 0.2509, 2.9173, -0.1308, 0.9136, 0.0706, -0.9477, -2.4055, 0.7462, 0.9538, -0.8179, -0.4440, -1.1852, -0.7789, -0.5011, -1.8987, -0.3188, 2.0584, 0.6024, -3.1321, -0.6272, -0.5745, 0.4365,
0.9436, -0.6201, 1.3551, -0.4047, 1.5559, 1.0733, -0.3982, -1.4490, -0.0592, 2.8637, 0.9025, -0.7550, 1.5398, -0.5594, 0.4108, -0.5978, 0.0613, -0.9095, -0.4517, -0.9100, -0.3021, -0.3001, -0.2229, 0.9803, 1.8816, 2.9271,
3.3045, -1.2201, -0.9750, 0.0013, -0.9089, -0.2484, 0.0234, -1.1106, -0.8048, 1.0323, -0.5966, 6.1361, -0.1659, -0.0031, 0.0573, -2.0689, -1.6038, 0.6123, 0.4766, -0.1283, -0.3109, -6.5703, 3.0460, -0.4466, -0.3737, -0.1055,



0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10
−5
0
5
10



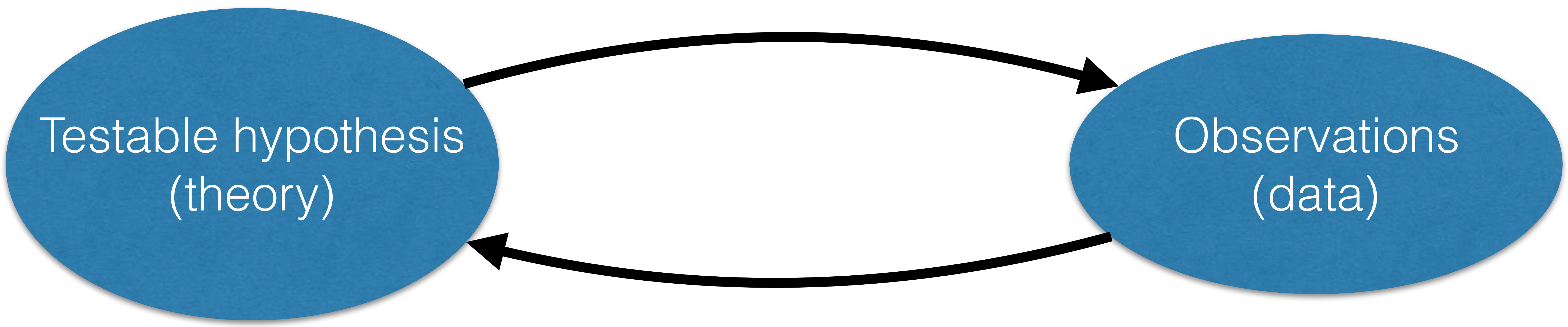
$\mu = -0.038; \sigma = 1.15$
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10
−5
0
5
10


0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10
−5
0
5
10

Data modelling in the scientific method

Fig. adapted from
Gregory (2005)

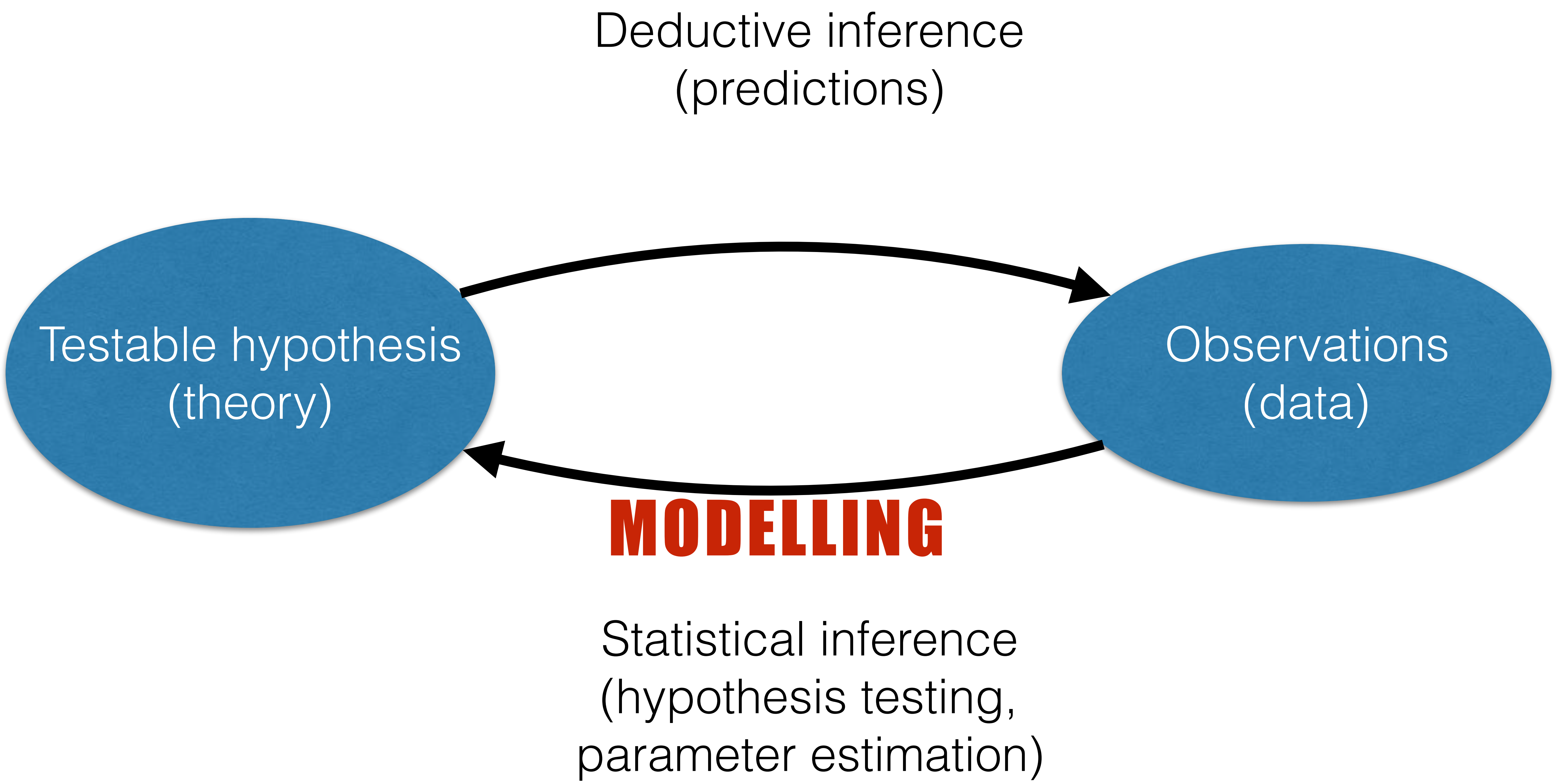
Deductive inference
(predictions)



Statistical inference
(hypothesis testing,
parameter estimation)

Data modelling in the scientific method

Fig. adapted from Gregory (2005)



Agenda para hoy

- **Introducción a los modelos lineales para la regresión.**
  - ¿Qué son los modelos lineales?
  - ¿Qué modelan?
  - ¿Cómo encuentro los valores de los parámetros?
  - …

- **Diagnóstico de modelos**
  - Residuos.
  - Palanca.
  - …

Modelling

The very basic idea

ICAS
icas.unsam.edu.ar

$$\underbrace{t_i}_{\text{data}} = \underbrace{m_i}_{\text{model}} + \underbrace{\epsilon_i}_{\text{error}} \qquad i = \{1, \ldots, N\}$$

$$m_i = f(X_i | \boldsymbol{\omega})$$

Modelling

The very basic idea

Target values

$$\underbrace{t_i}_{\text{data}} = \underbrace{m_i}_{\text{model}} + \underbrace{\epsilon_i}_{\text{error}} \qquad i = \{1, \ldots, N\}$$

$$m_i = f(X_i|\omega)$$

Modelling

The very basic idea

Target values

$$\underset{\textbf{data}}{t_i} = \underset{\textbf{model}}{m_i} + \underset{\textbf{error}}{\epsilon_i} \qquad i = \{1, \ldots, N\}$$

$$m_i = f(X_i | \omega)$$

Model parameters

icas.unsam.edu.ar

Modelling

The very basic idea

Target values

$$\underbrace{t_i}_{\textbf{data}} = \underbrace{m_i}_{\textbf{model}} + \underbrace{\epsilon_i}_{\textbf{error}} \qquad i = \{1, \ldots, N\}$$

$$m_i = f(X_i | \omega)$$

Features

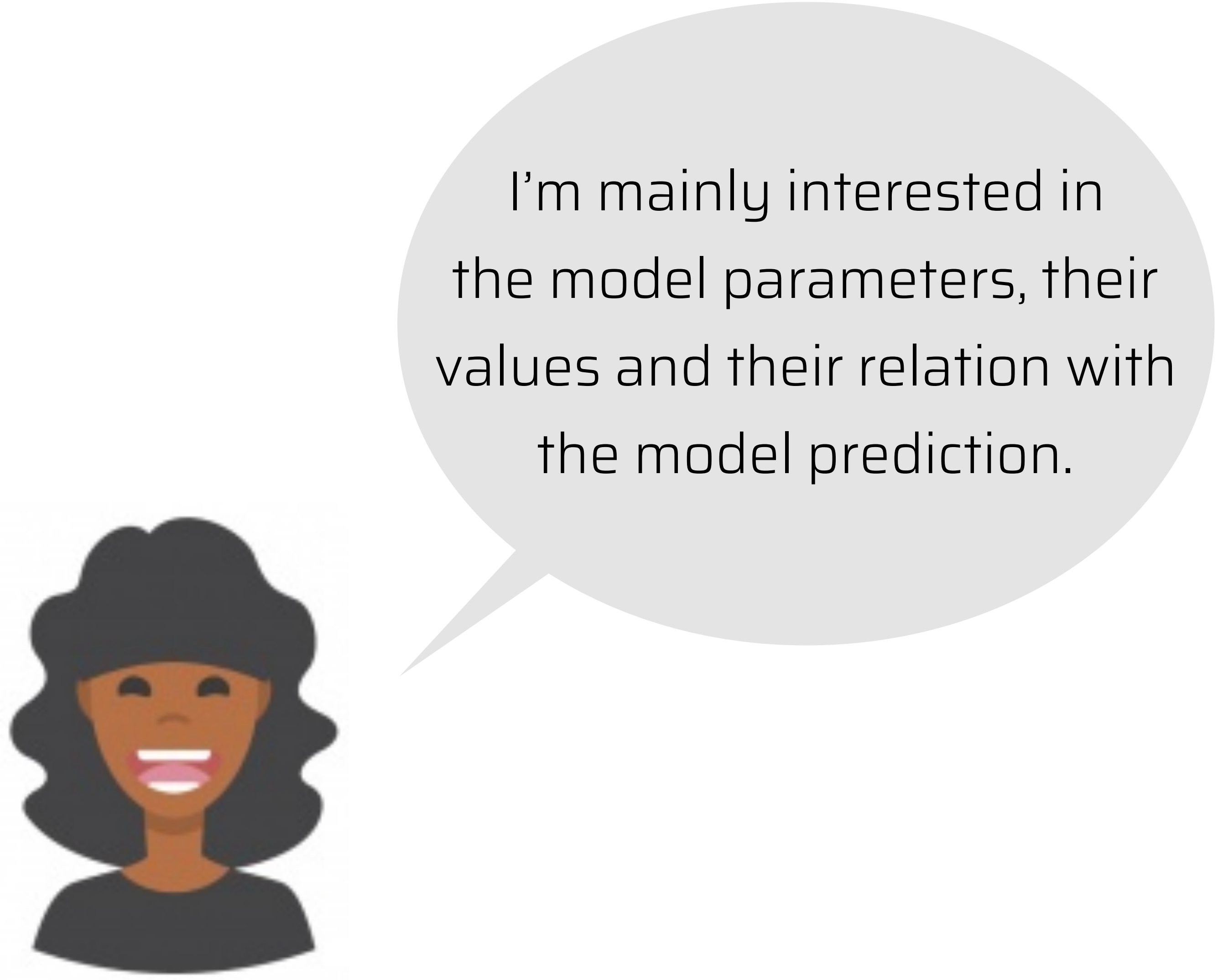
Model parameters

Modelling

The very basic idea

Target values

$$\underbrace{t_i}_{\text{data}} = \underbrace{m_i}_{\text{model}} + \underbrace{\epsilon_i}_{\text{error}} \qquad i = \{1, \ldots, N\}$$



$$m_i = f(X_i|\omega)$$

Features

Model parameters

Modelling

The very basic idea

Target values

$$t_i = m_i + \epsilon_i$$

data model error

$$i = \{1, \dots, N\}$$

I'm mainly interested in the model parameters, their values and their relation with the model prediction.

Alice

$$m_i = f(X_i | \omega)$$

Features

Model parameters

I want to create a model that gives the most accurate predictions for *unseen* values of the features X

Bob

Modelling

A probabilisitic view of linear regression

$$t_i = m_i + \epsilon_i$$

Modelling

A probabilisitic view of linear regression

$$t_i = m_i + \epsilon_i$$

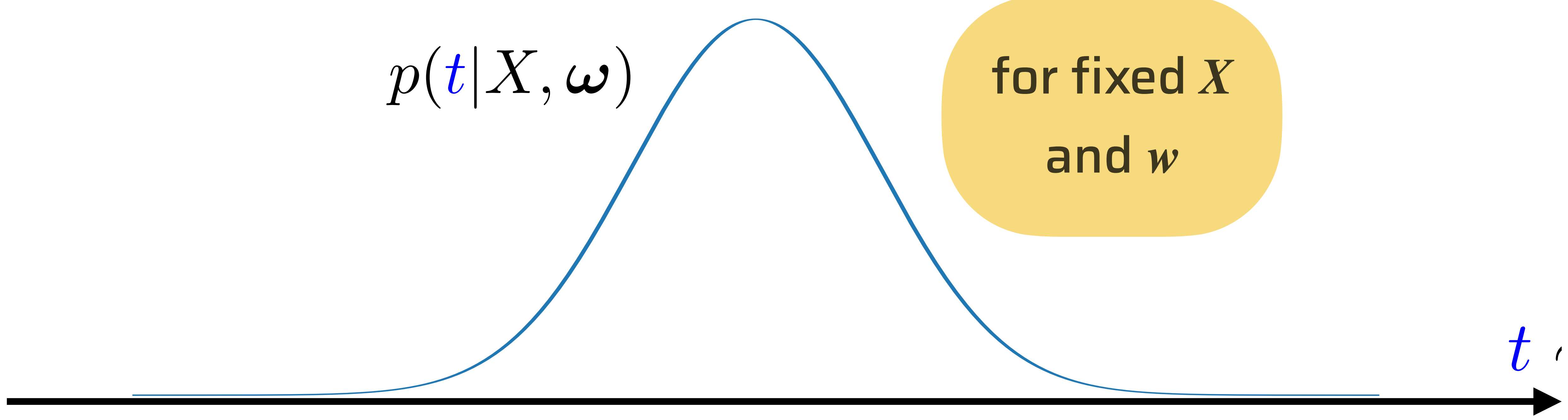
$$t \sim p(t|X, \boldsymbol{\omega})$$

Modelling

A probabilisitic view of linear regression

$$t_i = m_i + \epsilon_i$$

$$t \sim p(t|X, \boldsymbol{\omega})$$



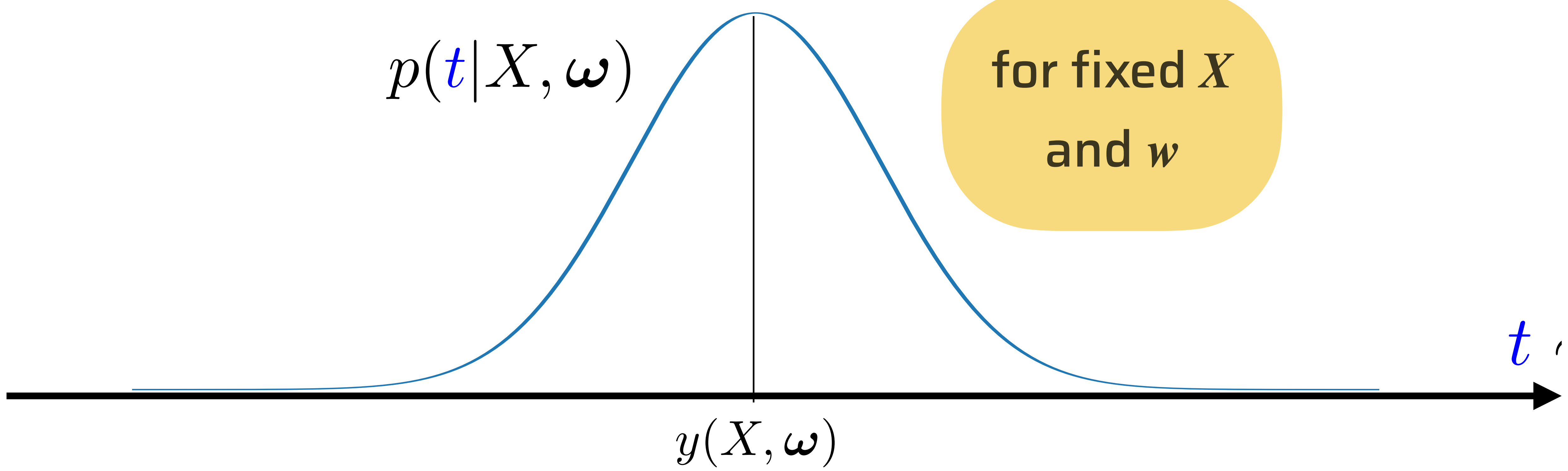
Modelling

A probabilisitic view of linear regression

$$t_i = m_i + \epsilon_i$$

$$t \sim p(t|X, \boldsymbol{\omega})$$

$$\mathbb{E}(t|X) = y(X, \boldsymbol{\omega})$$



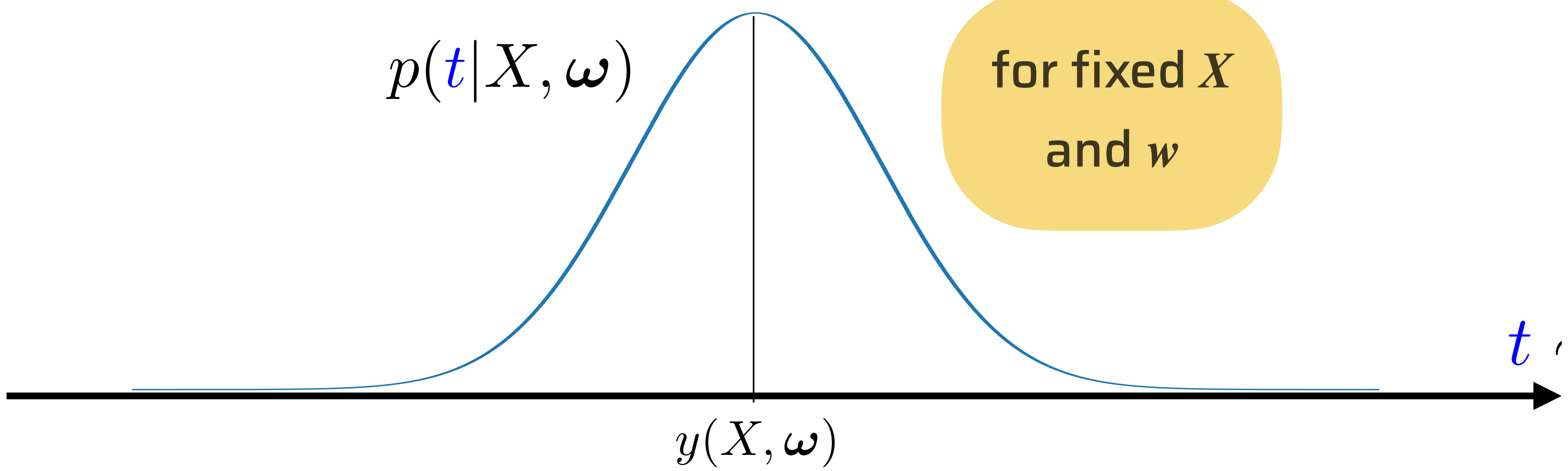
for fixed X and w

Modelling

A probabilisitic view of linear regression

$$t_i = m_i + \epsilon_i$$

$$t \sim p(t|X, \boldsymbol{\omega})$$



for fixed X and w

$$\mathbb{E}(t|X) = y(X, \boldsymbol{\omega}) = \omega_0 + \omega_1 \cdot X$$

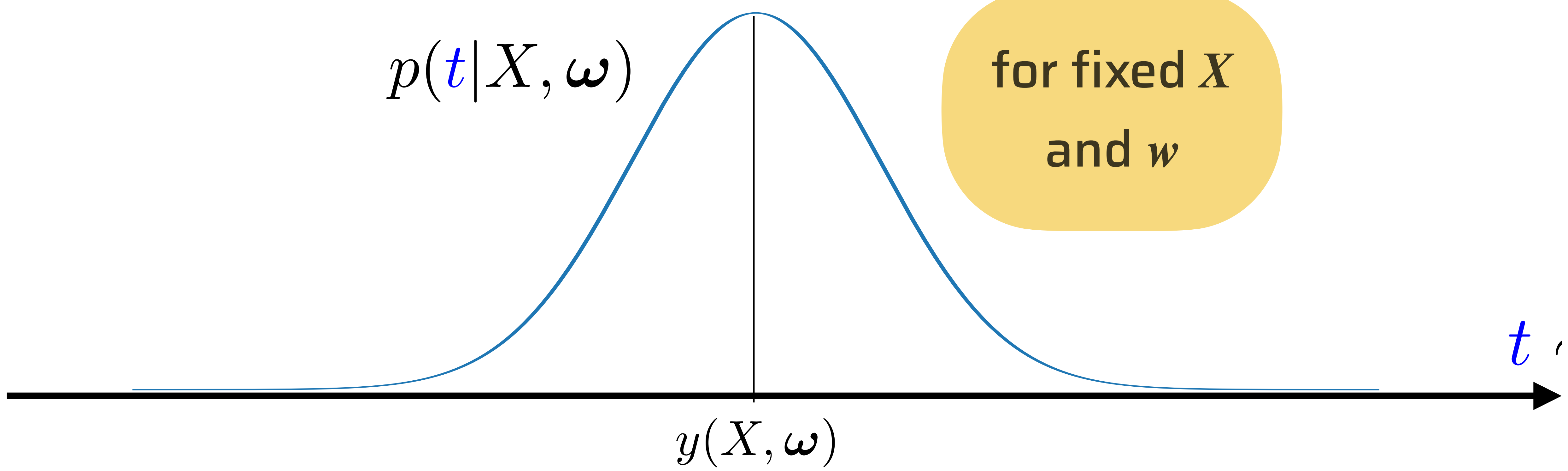
Simple Linear Regression

Modelling

A probabilisitic view of linear regression

$$t_i = m_i + \epsilon_i$$

$$t \sim p(t|X, \boldsymbol{\omega})$$



for fixed X and w

$$\mathbb{E}(t|X) = y(X, \boldsymbol{\omega}) = \omega_0 + \omega_1 \cdot X$$

Simple Linear Regression

$$y(\boldsymbol{X}, \boldsymbol{\omega}) = \omega_0 + \omega_1 \cdot X_1 + \omega_2 \cdot X_2 + \ldots + \omega_D \cdot X_D$$

Multiple Linear Regression

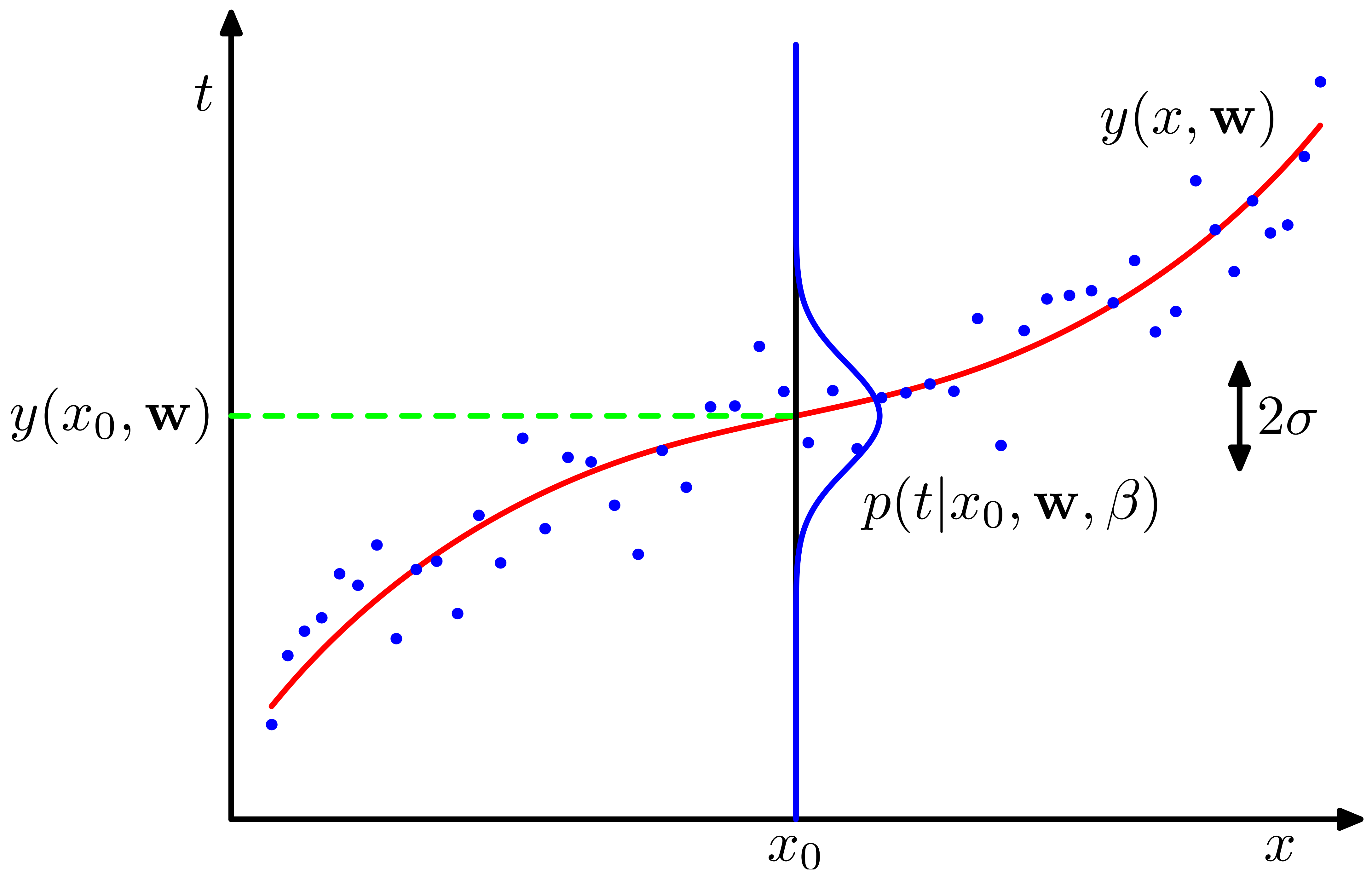
Modelling

Simple linear model - what are we actually modelling?

icas.unsam.edu.ar

$$t \sim N(y(X, \boldsymbol{\omega}), \sigma^2)$$

$$y = \omega_0 + \omega_1 \cdot x$$



figures from Bishop

Modelling

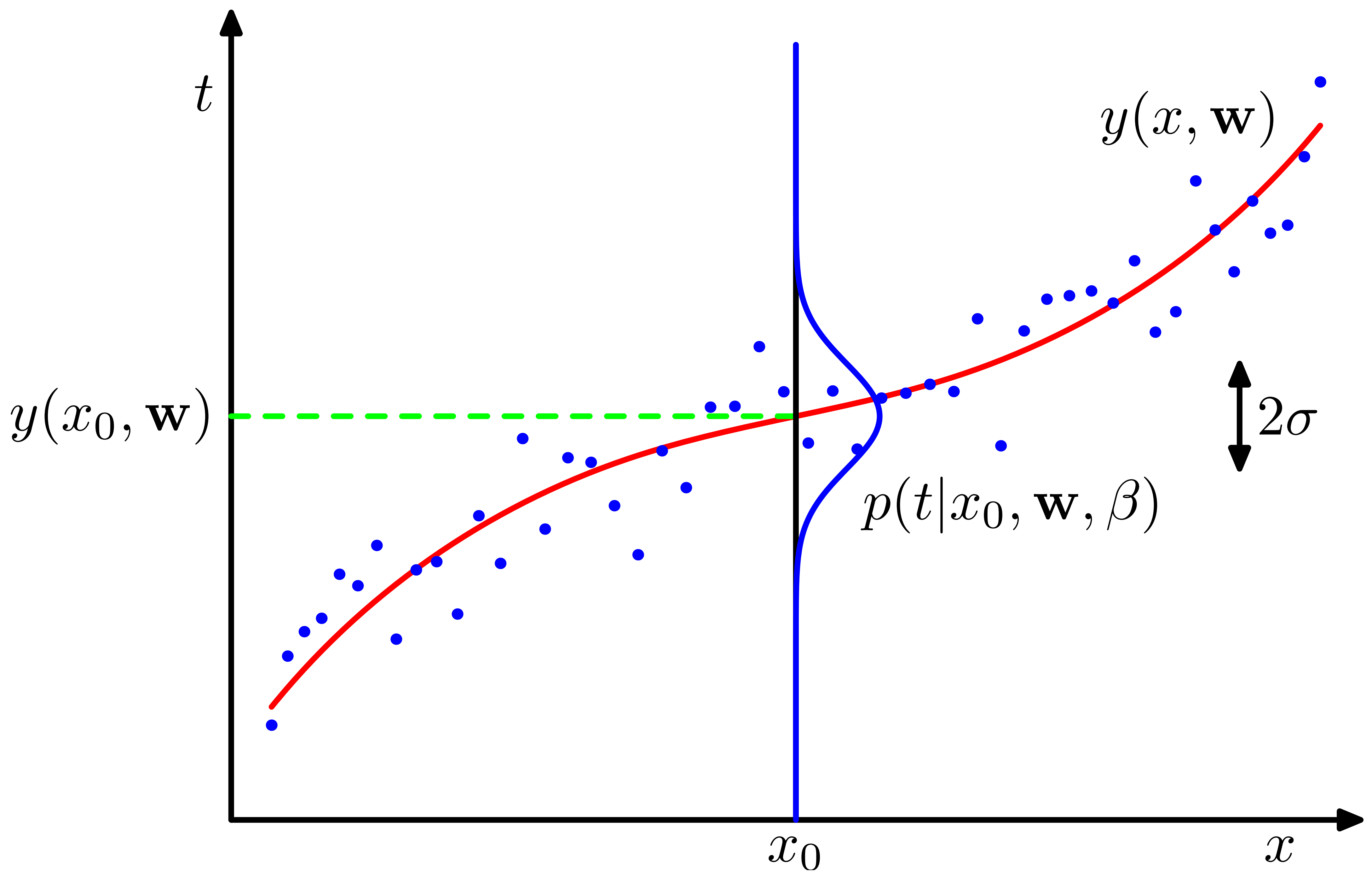
Simple linear model - what are we actually modelling?

$$t \sim N(y(X, \boldsymbol{\omega}), \sigma^2)$$

$$y = \omega_0 + \omega_1 \cdot x$$

We are assuming:

- The errors have zero mean $\mathbb{E}(\epsilon) = 0$
- Their distribution is Normal.
- They share a same covariance σ^2
- Errors are independent

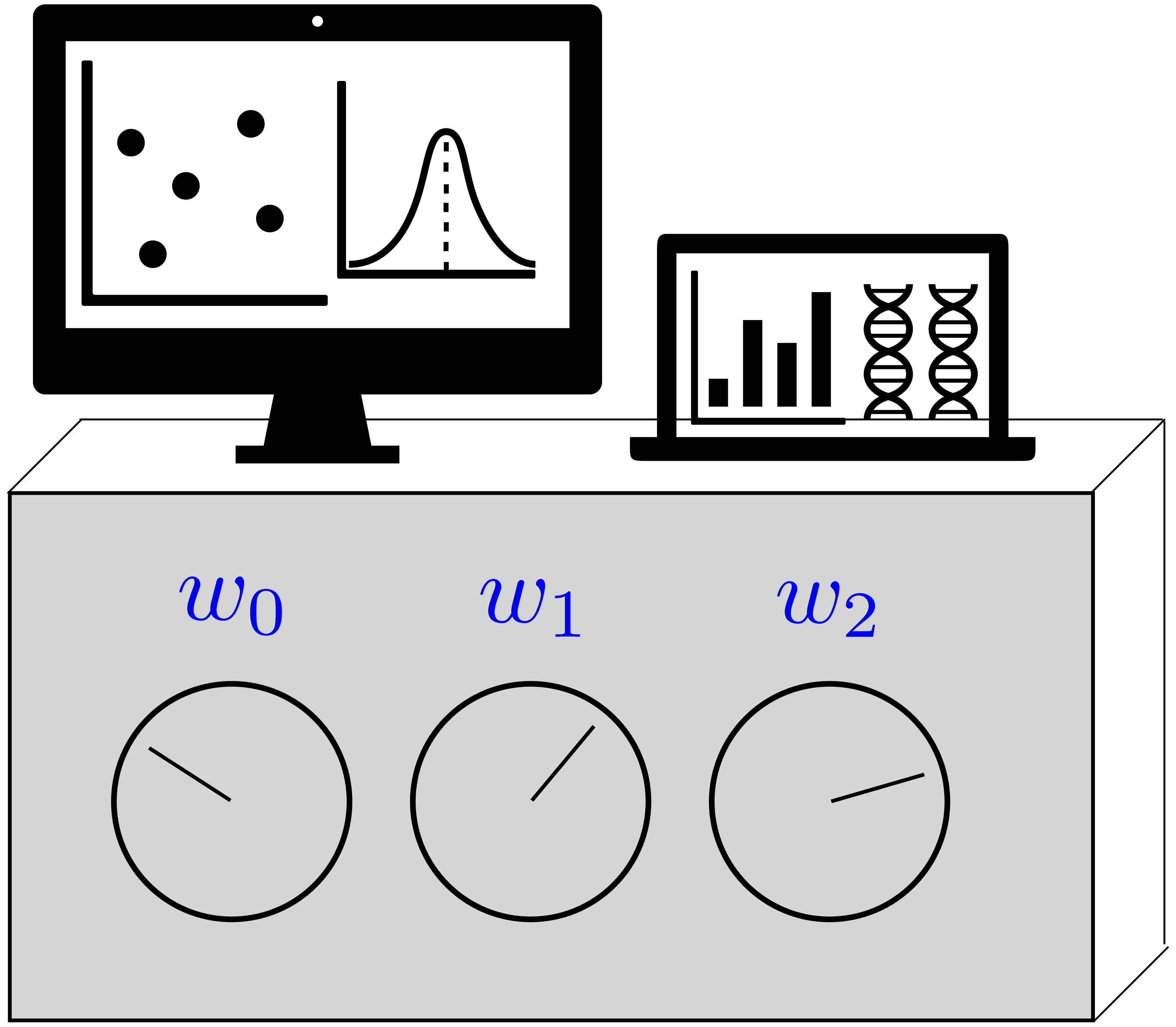


figures from Bishop

Model training

Adjusting the knobs

ICAS
icas.unsam.edu.ar

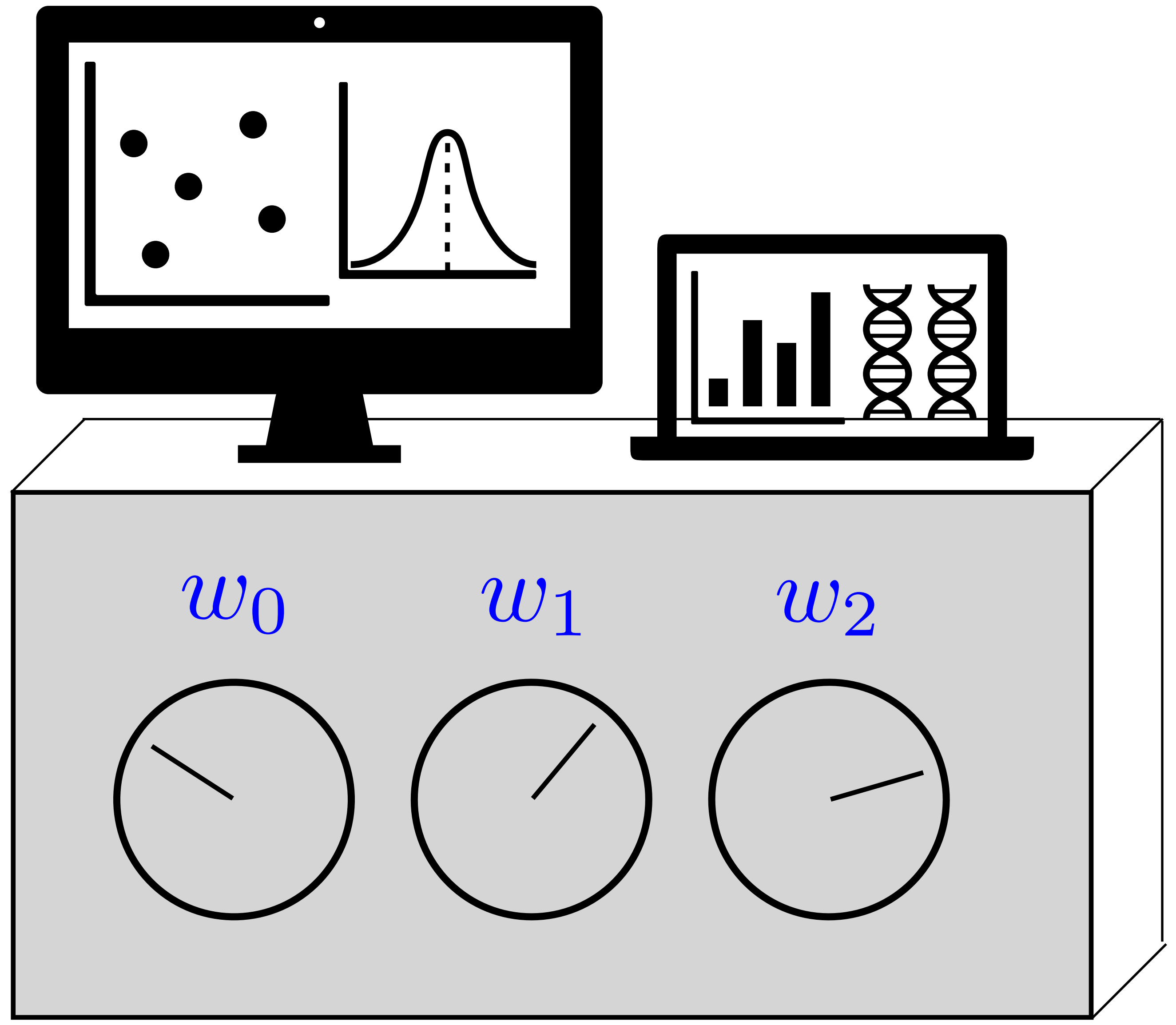


Model training

Adjusting the knobs

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}(\text{prediction}_i - \text{target}_i)^2$$

Mean Squared Error



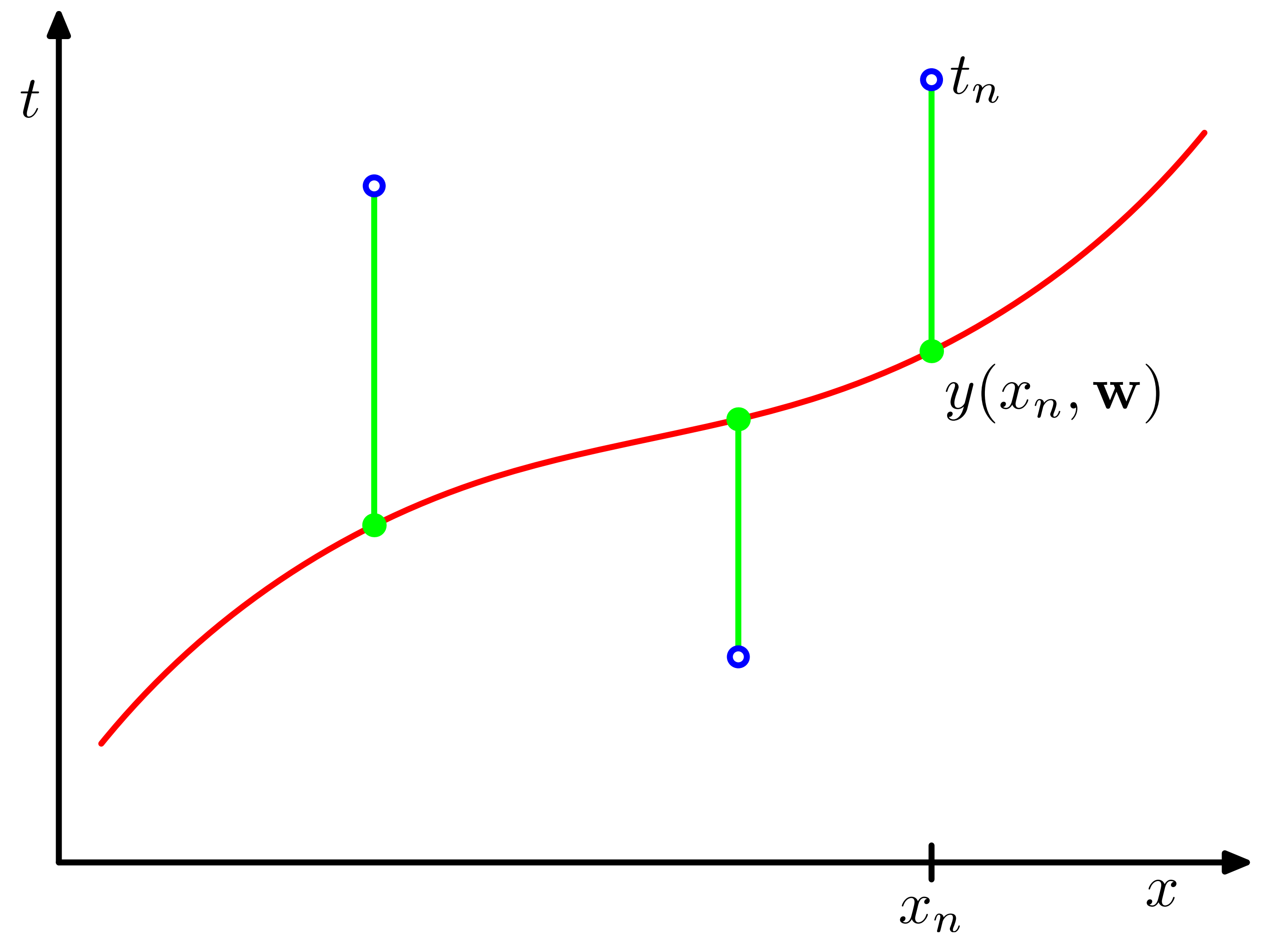


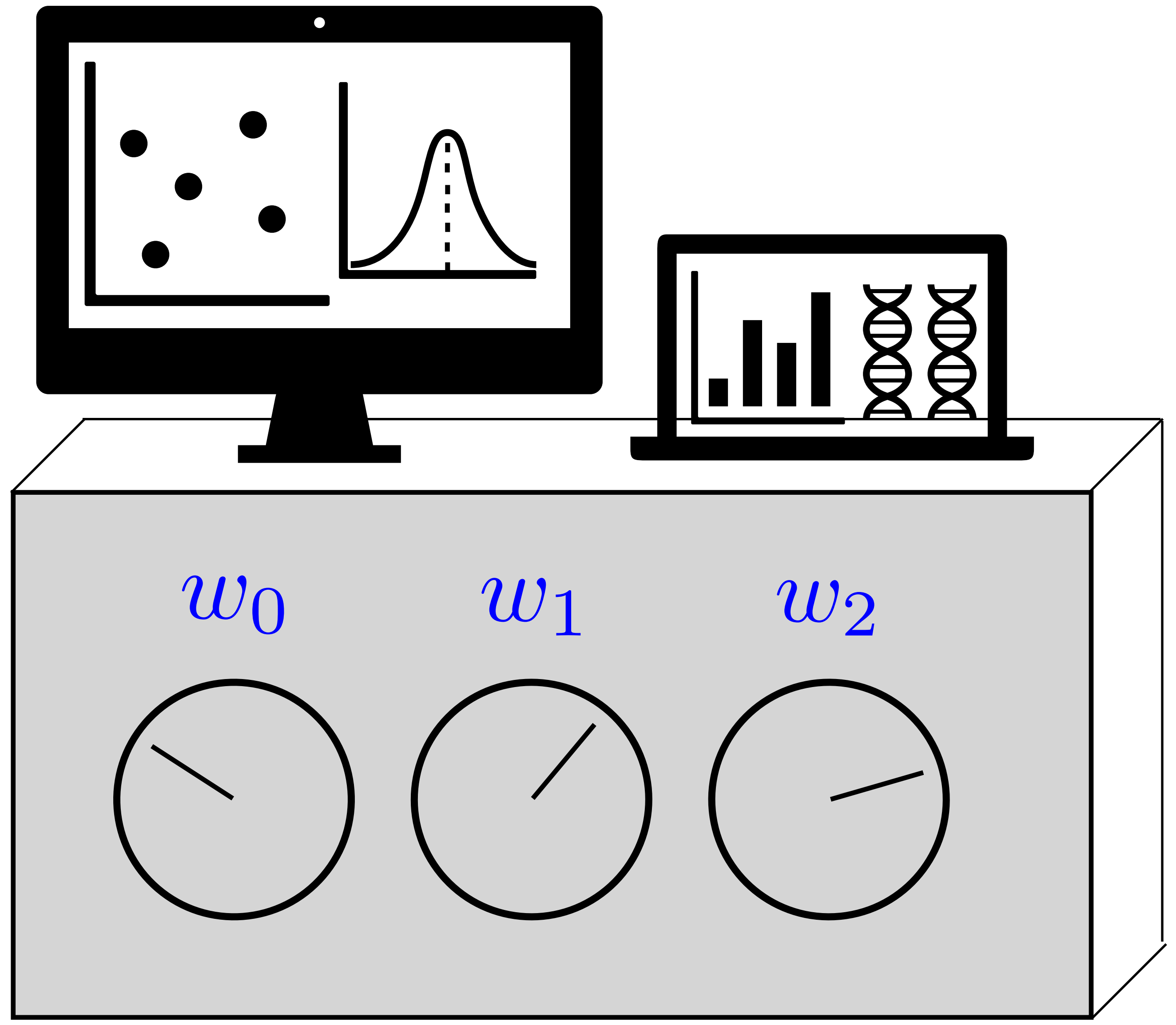
figure from Bishop

Model training

Adjusting the knobs

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}(\text{prediction}_i - \text{target}_i)^2$$

Mean Squared Error



$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2$$

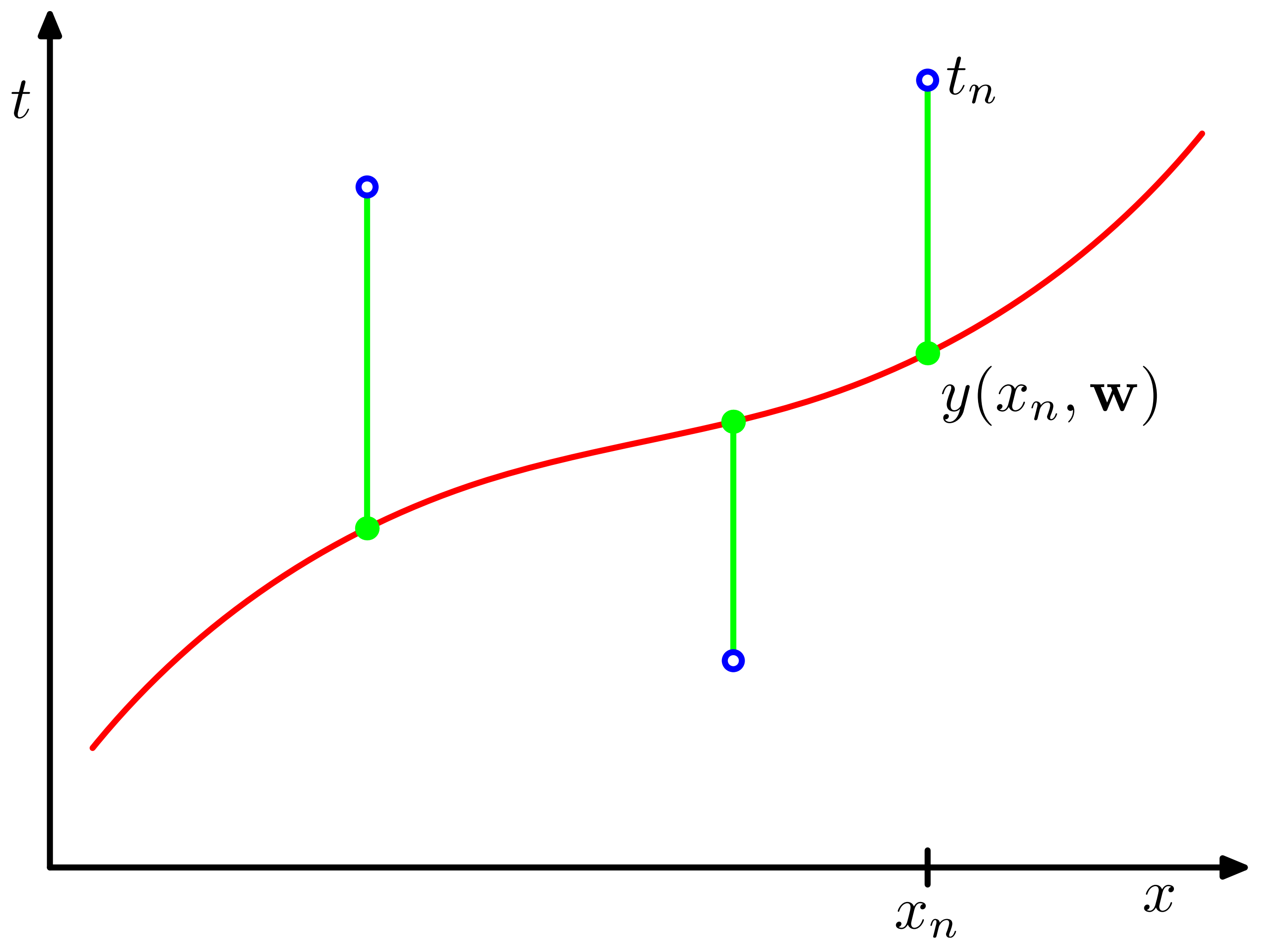


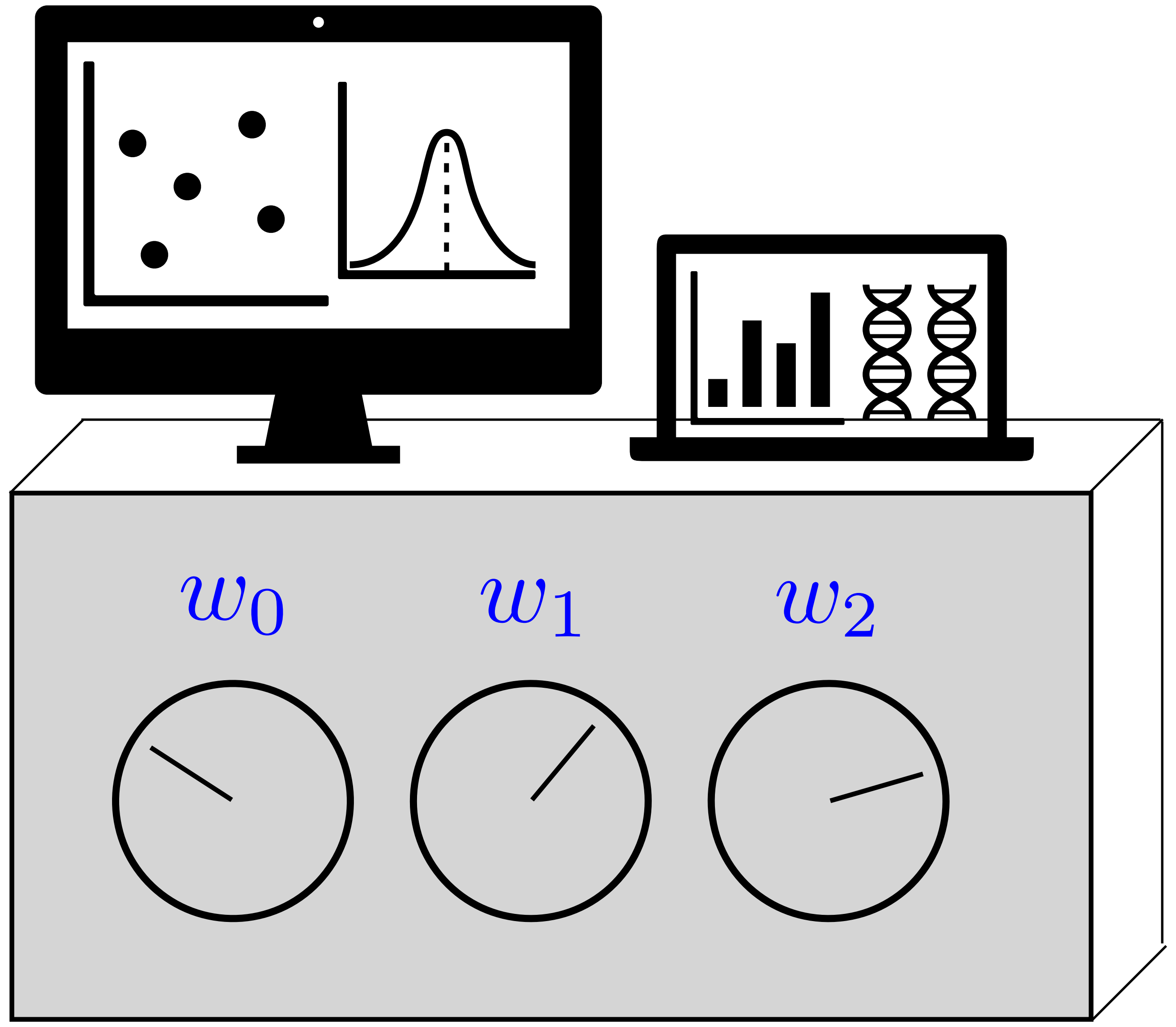
figure from Bishop

Model training

Adjusting the knobs

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}(\text{prediction}_i - \text{target}_i)^2$$

Mean Squared Error



$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2}$$

Root Mean Squared Error

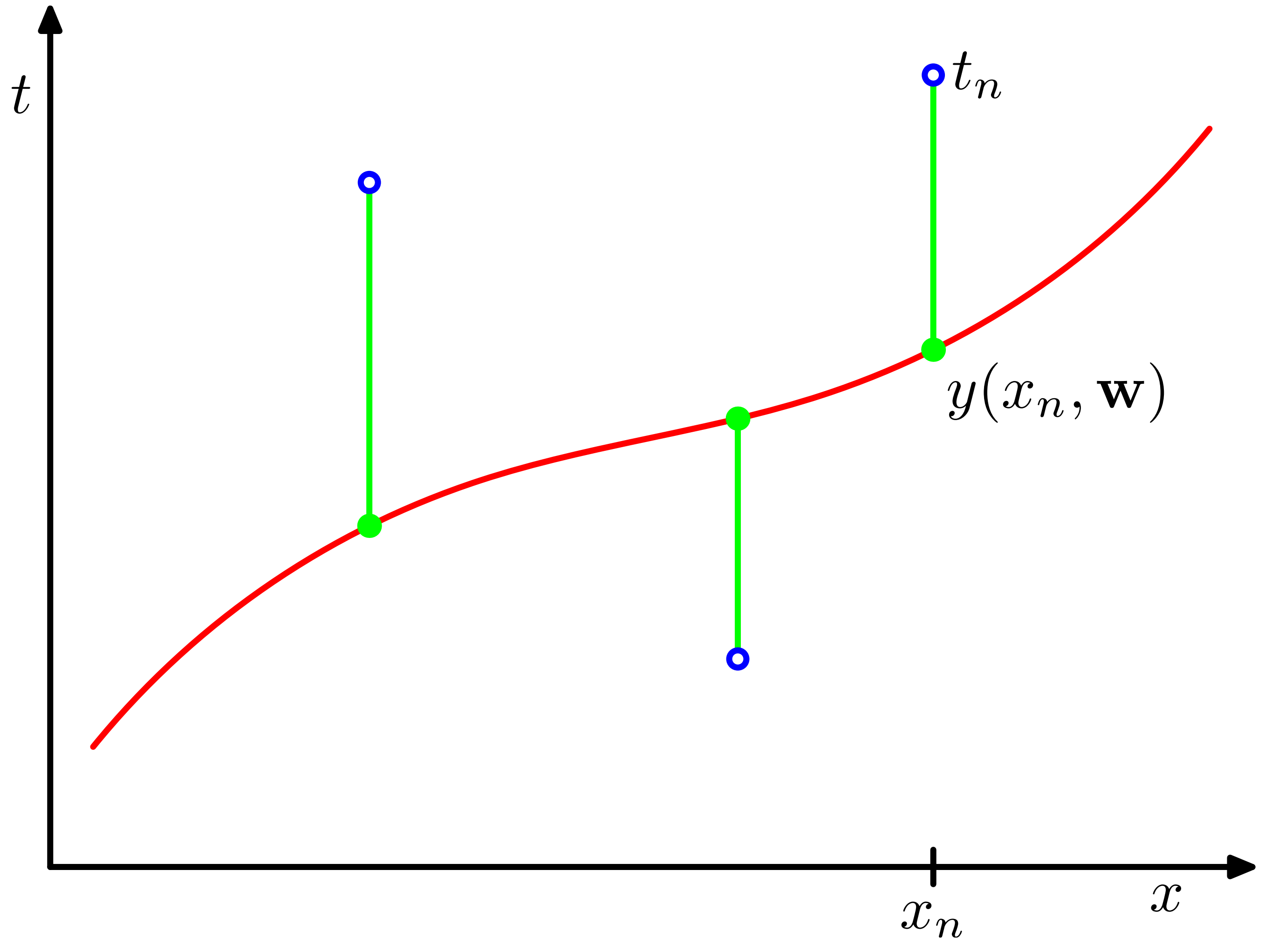


figure from Bishop

Modelling

Simple linear model - normal equations

$$y = \omega_0 + \omega_1 \cdot x \qquad MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2$$

$$\boxed{\begin{aligned} \textstyle\sum_{i=1}^{N} \left[t_i - (\hat{\omega_0} + \hat{\omega_1} \cdot x_i)\right] &= 0 \\ \textstyle\sum_{i=1}^{N} \left[t_i - (\hat{\omega_0} + \hat{\omega_1} \cdot x_i)\right] x_i &= 0 \end{aligned}}$$

Modelling

Simple linear model - normal equations

$$y = \omega_0 + \omega_1 \cdot x \qquad MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2$$

$$\boxed{\begin{aligned} \textstyle\sum_{i=1}^{N}\left[t_i - (\hat{\omega_0} + \hat{\omega_1} \cdot x_i)\right] &= 0 \\ \textstyle\sum_{i=1}^{N}\left[t_i - (\hat{\omega_0} + \hat{\omega_1} \cdot x_i)\right] x_i &= 0 \end{aligned}}$$

Normal equations

Modelling

Simple linear model - normal equations

$$y = \omega_0 + \omega_1 \cdot x \qquad MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2$$

$$\boxed{\begin{aligned} \textstyle\sum_{i=1}^{N}\left[t_i - (\hat{\omega_0} + \hat{\omega_1} \cdot x_i)\right] &= 0 \\ \textstyle\sum_{i=1}^{N}\left[t_i - (\hat{\omega_0} + \hat{\omega_1} \cdot x_i)\right] x_i &= 0 \end{aligned}}$$

Normal equations

$$\begin{aligned} \hat{\omega_1} &= \textstyle\sum_{i=0}^{N}(x_i - \bar{X})(y_i - \bar{Y}) \left[\sum_{i=0}^{N}(x_i - \bar{X})^2\right]^{-1} \\ \hat{\omega_0} &= \bar{Y} - \hat{\omega_1}\bar{X} \end{aligned} \qquad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \ .$$

Model extension

Multiple linear model

Simple Linear Regression

$$y(x, w_0, w_1) = w_0 + w_1 x \quad .$$

Multiple Linear Regression

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \cdots + w_D x_D \quad .$$

Model extension

Multiple linear model

Simple Linear Regression

$$y(x, w_0, w_1) = w_0 + w_1 x \quad .$$

Multiple Linear Regression

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \cdots + w_D x_D \quad .$$

More generally:

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{i=1}^{D} w_j \phi_j(\boldsymbol{x_i}) = \sum_{j=0}^{D} w_j \phi_j(\boldsymbol{x_i}) = \boldsymbol{w}^T \boldsymbol{\phi}_i$$

Multiple linear model

Matrix Notation

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{i=1}^{D} w_j \phi_j(\boldsymbol{x_i}) = \sum_{j=0}^{D} w_j \phi_j(\boldsymbol{x_i}) = \boldsymbol{w}^T \boldsymbol{\phi}_i$$

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{\phi}_i \qquad i = \{1, \ldots, N\}$$

Multiple linear model

Matrix Notation

icas.unsam.edu.ar

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{i=1}^{D} w_j \phi_j(\boldsymbol{x_i}) = \sum_{j=0}^{D} w_j \phi_j(\boldsymbol{x_i}) = \boldsymbol{w}^T \boldsymbol{\phi}_i$$

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{\phi}_i \qquad i = \{1, \ldots, N\}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \boldsymbol{\Phi} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix}$$

(N x 1) = (N x D) (D x 1)

Multiple linear model

Matrix Notation

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{i=1}^{D} w_j \phi_j(\boldsymbol{x_i}) = \sum_{j=0}^{D} w_j \phi_j(\boldsymbol{x_i}) = \boldsymbol{w}^T \boldsymbol{\phi}_i$$

$$y_i(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{\phi}_i \qquad i = \{1, \ldots, N\}$$

Design Matrix

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \boldsymbol{\Phi} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix} \qquad \boldsymbol{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$$

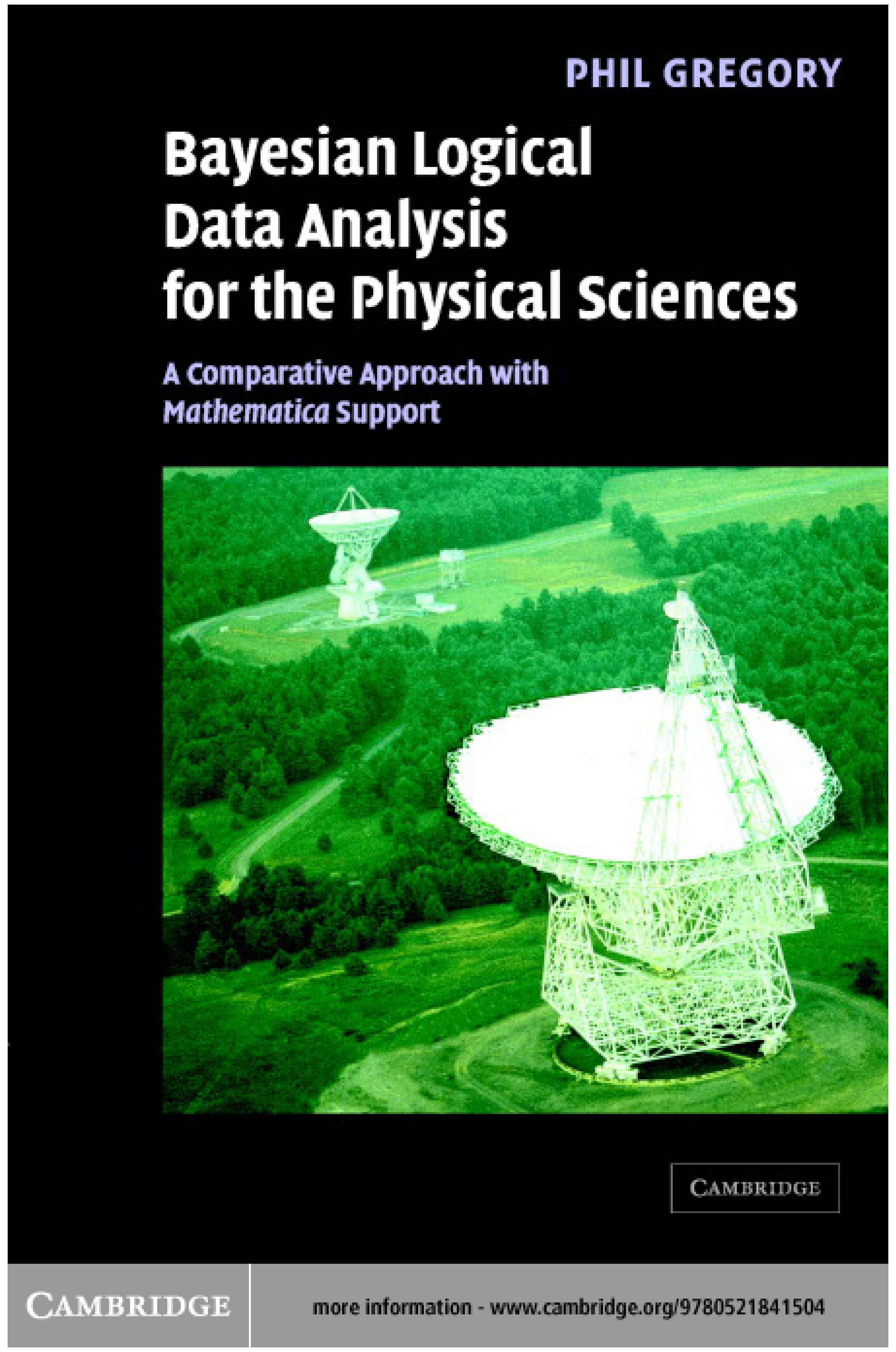
(N x 1) = (N x D) (D x 1)

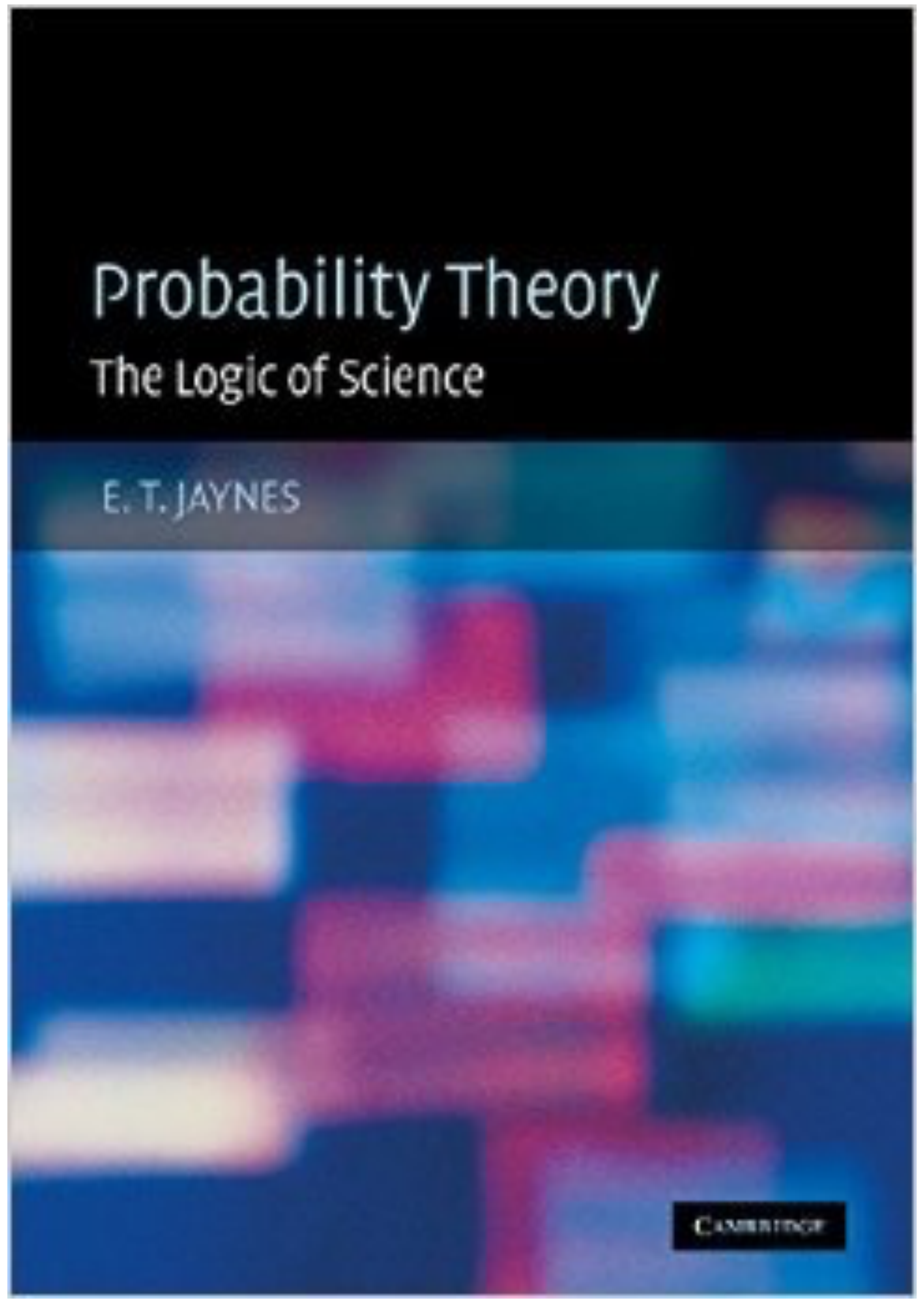
Normal Equations

The normal equations can also be solved analytically in this case.

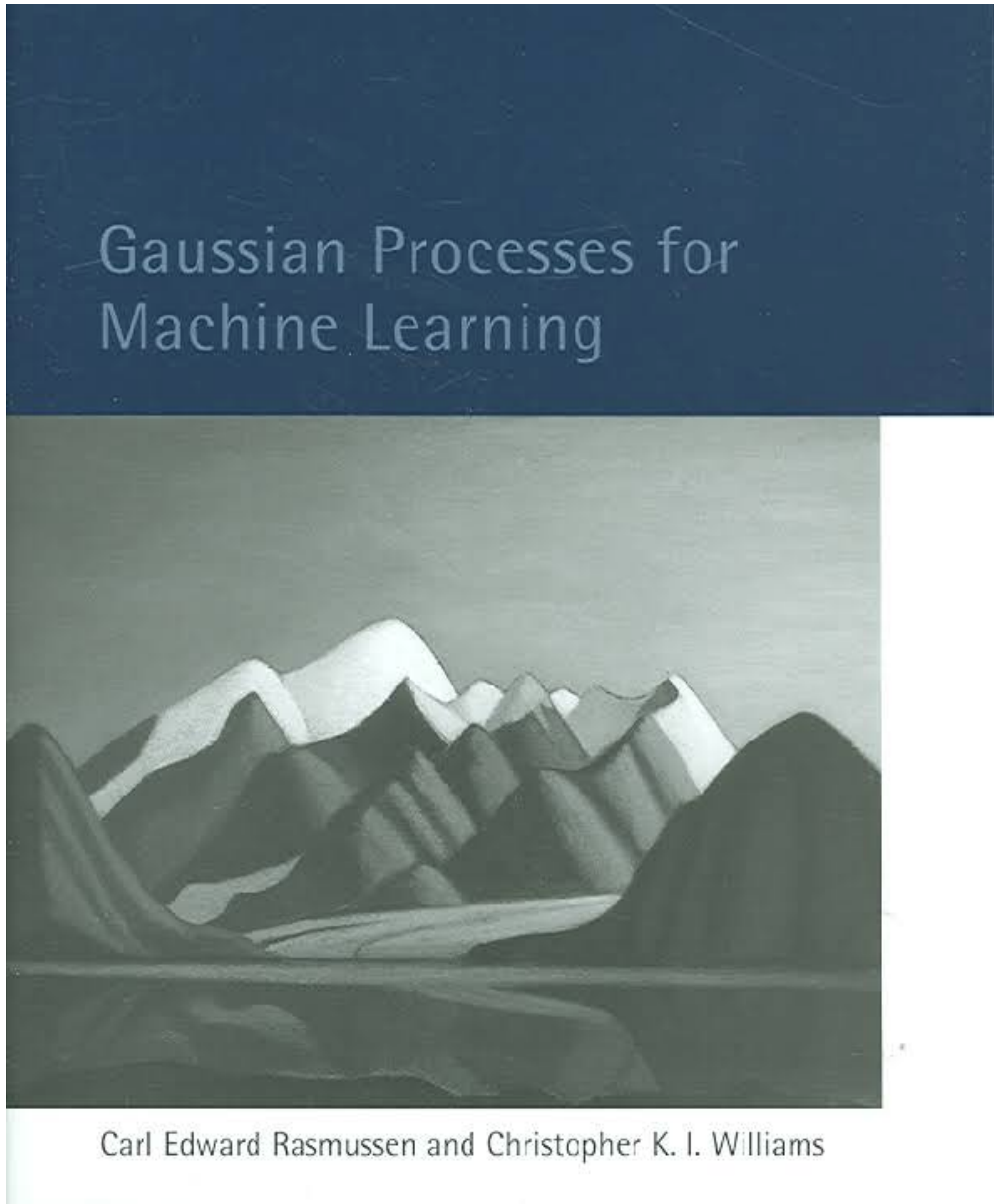
$$\boxed{\mathbf{w}_{\mathrm{ML}} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}}$$

Main references - Bayesian





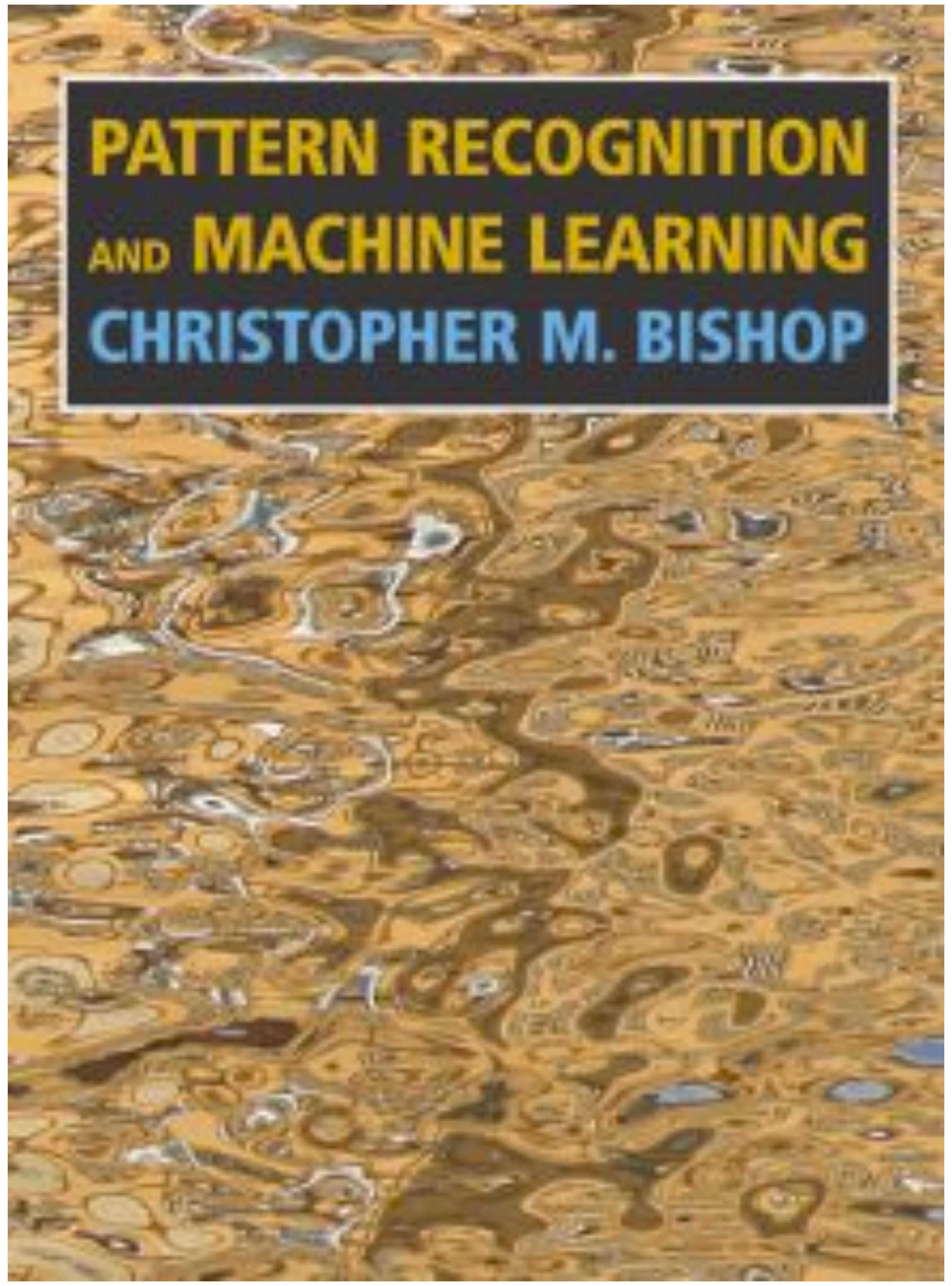
Main references - Gaussian processes



Available online

`http://www.gaussianprocess.org/gpml/`

Main references - Machine Learning



Available online